

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2002, 01:33:20 ; Search time 1741.88 Seconds
(without alignments)
17011.491 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gccaccatcgctgcgcctnt.....aaaaaaaaaaaaaaaaaaaaa 1416

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_ov:*
5: gb_om:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Query Score	Match length	ID	Description

1	1343	94.8	1385	9	AF069987	AF069987 Homo sapi
2	863.6	61.0	1365	10	BC021634	BC021634 Mus muscu
3	859.6	60.7	1338	10	AF069988	AF069988 Mus muscu
4	572.2	40.4	4079	9	AF069984	AF069984 Homo sapi
5	572.2	40.4	110000	2	AL554714_4	Continuation (5 of
6	572.2	40.4	110000	2	AL554714_5	Continuation (6 of
7	572.2	40.4	167863	2	AL590651	AL590651 Homo sapi
8	572.2	40.4	200882	9	AL591806	AL591806 Human DNA
9	356.2	25.2	1214	5	AF284575	AF284575 Xenopus l
10	326.8	23.1	4481	10	AF069985	AF069985 Mus muscu
11	326.8	23.1	211771	2	AC084821	AC084821 Mus muscu
12	326.8	23.1	213203	2	AC087229	AC087229 Mus muscu
13	308.4	21.8	178418	2	AC105589	AC105589 Rattus no
14	280	19.8	441	11	G13436	G13436 human STS W
15	276.2	19.5	847	11	G72919	G72919 MARC 2849-2
16	182.8	12.9	1521	3	AF069989	AF069989 Drosophill
17	182.8	12.9	131853	2	AC017755	AC017755 Drosophill
18	182.8	12.9	168250	3	AC093121	AC093121 Drosophill
19	182.8	12.9	173494	3	AC093501	AC093501 Drosophill
20	182.8	12.9	174279	2	AC006169	AC006169 Drosophill
21	182.8	12.9	190351	3	AC105293	AC105293 Drosophill
22	182.8	12.9	298640	3	AE003467	AE003467 Drosophill
23	168.4	11.9	1385	3	AF069986	AF069986 Caenorhab
24	160.8	11.4	34013	2	AC010013	AC010013 Drosophill
25	144.2	10.2	1172	8	AF372904	AF372904 Arabidops
26	113.8	8.0	1063	10	BC020153	BC020153 Mus muscu
27	113.8	8.0	1292	10	AF284573	AF284573 Mus muscu
28	112.8	8.0	4762	1	AB017194	AB017194 Plectonem
29	112.6	8.0	25350	8	SPBC651	AL035570 S.pombe c
30	112.4	7.9	10466	1	AE004861	AE004861 Pseudomon
31	106.4	7.5	110908	1	SYCSLRD	D64002 Synchocyst
32	105.4	7.4	235050	1	AL414158	AL414158 yeastnia
33	105.2	7.4	203050	1	AL646071	AL646071 Ralstonia
34	104.4	7.4	169364	3	CEY56A3A	AL132860 Caenorhab
35	102.6	7.2	343550	1	AP003587	AP003587 Nostoc sp
36	102	7.2	962	9	AF284574	AF284574 Homo sapi
37	102	7.2	1019	9	AF260334	AF260334 Homo sapi
38	101.4	7.2	10765	1	AE008331	AE008331 Agrobacte
39	101.4	7.2	11689	1	AE009281	AE009281 Agrobacte
40	101.2	7.1	57093	2	AC012711	AC012711 Drosophill
41	101.2	7.1	148847	3	AC008234	AC008234 Drosophill
42	101.2	7.1	175053	3	AC010581	AC010581 Drosophill
43	101.2	7.1	224400	3	AE003682	AE003682 Drosophill
44	100.4	7.1	965	9	BC020620	BC020620 Homo sapi
45	99.8	7.0	12748	1	AE005759	AE005759 Caulobact

ALIGNMENTS

RESULT	1	AF069987	1385 bp	mrna	linear	PRI 23-JUL-1998
LOCUS	AF069987	Homo sapiens	nitrilase 1 (NIT1)	mrna, complete cds.		
DEFINITION	AF069987					
ACCESSION	AF069987					
VERSION	AF069987.1	GI:3228665				
KEYWORDS						
SOURCE						
ORGANISM						
		human.				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE		1 (bases 1 to 1385)				
AUTHORS		Pekarsky,Y., Campigillo,M., Siprashvili,Z., Druck,T., Sedkov,Y.,				
		Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,				
		Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.				
		Nitrilase and phl homologs are encoded as fusion proteins in				
		drosophila melanogaster and caenorhabditis elegans				
		Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)				
JOURNAL		98337986				
MEDLINE		2 (bases 1 to 1385)				
REFERENCE		Pekarsky,Y., Campigillo,M., Siprashvili,Z., Druck,T., Sedkov,Y.,				
AUTHORS		Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,				
		Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.				

REMARK
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRK Plate: 18 Row: e Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6754855.

FEATURES

source
1..1365
/organism="Mus musculus"
/db_xref="locusID:27045"
/db_xref="taxon:10090"
/clone="MGC:13825 IMAGE:4008543"
/issue.type="Mammary tumor metastasized to lung.
MMTV-LTR/Int1 model. Expression driven by an MMTV-LTR
enhancer."
/clone.lib="NCI-CGAP-Lu30"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
59..1030
/codon_start=1
/product="nitrilase 1"
/protein_id="AAH21634.1"
/db_xref="GI:18204913"
/translation="MLGFTPRPHOLCTGYRLRTPVLTOTPRPRMTSSSTWELPL
VAVCOYTPNKQENKCAELVQEARACAFPEAPDFIARBPATLLISPLN
GVLLGYTSOLAREBCGIMSLGSHFHEKQMDONOKLYNCHVLINSQSVASAKRTHL
CVELIPGQPKRESNTKPGKTLPEPKTPAGVGLATYDMFEPISLKLAAQAEI
LYSSAFGSVTPAHWEVLRLARALISQCVIAAAGCGHNEFRASGYSMVDPWGT
VVARSEGGGLDLHLFILOMRHOLFVFRREDLVGSLHPLS"

CDS
368 a 334 c 331 g 332 t
BASE COUNT
ORIGIN

Query Match 61.0%; Score 863.6; DB 10; Length 1365;
Best Local Similarity 82.1%; Pired. No. 9.4e-244;
Matches 1076; Conservative 0; Mismatches 204; Indels 30; Gaps 6;

QY 110 gctgggcttaccacagcctctccacagatctcctctgtccctctgtctcgtgactccg 169
DB 61 GCTGGGCTTCATCACCAGGCTCTCTCAACAA-----CTCCTGTGTACCGGATACCG 111
QY 170 gataccacactcagactctgtgtgtcagccagagccagagagcattgactctc 229
DB 112 ATTACTTCGAAACCCCACTATTGTACTCAGCCAGGCCAGAACCAATGTC---CTATC 168
QY 230 ttccctctggaactcccccgtgtgtgtgtgtccaggtatacatcgacgcagaaca 289
DB 169 AACTTCTGTGAGCTGCGCCCTGGTGGCTGTGTGCGAGTAAACATCAACACCAACAAAGCA 228
QY 290 acagaacttaaacatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 349
DB 229 AGAGAACTTTAAACATGTCCTAGTGTGTTCAAGAGCTGCCAGACTGGGTCTTCCCT 288
QY 350 ggtctccctgagcagcattgactcattgacagcagcagcagcagcagcagcagcact 409
DB 289 GGCCTTTCCTGCTGAGGCACTTTGACTTATTTGACAGCAACACCTGCGCAGACTTACTCCT 348
QY 410 gtctgaaccactgt 469
DB 349 GTCCGAACCACTGAATGGGATCTTTTGGGCCAATATAGCAGCTTCCAGGGAATGTGG 408

QY 470 actctgctgtcctctgggtgtgtttccatgagcgtgtgccaagactggagagactcaga 529
DB 409 AATCTGCTGTCTTGGGCGGTTTCCAGAGAGCGGCCAAACATGGGACGAATACAAA 468
QY 530 aactacaatgtcagctgtcgtcgaacagcaagagggcagatgtgtgacattcaagaa 589
DB 469 AATCTACATTTGTATGTGCTTTTGAACAGCAAGGATCACTAGTGGCCATTACAGAA 528
QY 590 gacacatctgtgtgacgttagaagaltccagagcaggagcctatgtgtgaagcaactac 649
DB 529 GACACATCTGTGCGATGTAGAGATCCAGGTCAGGGGCCGATGAGAGAAACCACTATAC 588
QY 650 catgctgggcccagctcttgaatcactgttagaacaccagcaagatgtgtcagc 709
DB 589 CAACCTGTGAGGCACTCTTGAAGCACCCTGTAAAGACCGGCTGGCAAGTTGTATAC 648
QY 710 tgtctgtatgacatgtggtgtccctgaactcctctgtgcaattgtgtcaggtgagcaga 769
DB 649 AATCTGTATGACATGTGGGTTCCCTGAACCTTCTTTGAATTTGGCTCAAGCTGGGCGA 708
QY 770 gatacttaactatcccttcagcttctgtgattccattacaagcccaactgtgtgtgt 829
DB 709 AATCTTACTTATTTCTTCAAGCCTTTGGATGTGTACAGTCCAGCCACTGGAGGTGCT 768
QY 830 gctgagggcccggtgtctatcgaaaccaggtgtatgtatgtgtgtgtgtgtgtgtgt 889
DB 769 GCTGGGCGCCGCGCCCATGTCAATCTCAATGTATATACACACGCGCATGTGTGACG 828
QY 890 ccacacatgagaagagcaagttatgtgccaagcattgtgtgtgtgtgtgtgtgtgtgt 949
DB 829 CCACCATGAAGAAACAGACCAAGTATATGGCATTAGATGTGTGTACCCGTGGGCAAGT 888
QY 950 ggtggccgctgtctgt 1009
DB 889 GGTGGCCCGCTGTCCAGAGGAGCAGGCGCTGTGCTTGTGTAATGTATCTCCACTTCT 948
QY 1010 gctgacgt 1069
DB 949 ACAACAGATGGCGCAACACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1008
QY 1070 tctgt 1127
DB 1009 TCTGGGTATTCACCTCTTAAATGCTTTTAACTTCTGTGTGTGTGTGTGTGTGTGTGT 1067
QY 1128 ccccccactgt 1187
DB 1068 -----CAACTGGAGGTGATGCTACTGTGACTGTGAGCGACAGACACAAGCAGCG 1116
QY 1188 ctcccctcactgt 1246
DB 1117 TTCCCTTCTTCTGAGAACTTAACTCTTTTAAATGGAACACAGGCTTAAATCAAGTCCAT 1176
QY 1247 agaaacttcaactgt 1306
DB 1177 GGGAA-----AGAAAGACTTTTACCTGACCTGACATTTCAGTTTCAAGAAATGGGATTTT 1233
QY 1307 atatagtcattgttatttcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1366
DB 1234 ATATAGTACACTGTTTATTTCAATGAAGTGAAGTTATGCTGAGGCGTGTGAGCAGCACTGCG 1293
QY 1367 atgtgaaaaataataataataataataataataataataataataataataataata 1416
DB 1294 ATTGAAAAAAATTTATTTATTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1343

RESULT 3
AF069988 1338 bp mRNA linear ROD 23-JUL-1998
LOCUS AF069988
DEFINITION Mus musculus nitrilase 1 (Ntl1) mRNA, complete cds.
ACCESSION AF069988
VERSION AF069988.1 GI:3228667
KEYWORDS

[illegible]

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AF069984	4079 bp DNA	linear	PRI 23-JUL-1998		human.	Homo sapiens
LOCUS	Homo sapiens ntr1 gene, alternatively spliced	AF069984				
DEFINITION	product, complete cds.	AF069984				
ACCESSION	AF069984.1	GI:3242977				
VERSION						
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					

QY 869 ggcacacacagctgttgacgcccacacatgagagagaaagttatggccacagcatgt 928
|||||
Db 105614 GGCACACACAGCTGTGAGCGCCACCATGAGAGAGCAATTTATGGCCACGATGGT 105673
QY 929 ggtagacccctggggaacagctgtgcccgcctgtctctgaggggacagccctctgtc 988
|||||
Db 105674 GGTAGACCCCTGGGGAACAGTGTGGCGCGCTGTGAGGGGCGAGGCGCTGTGCTTGC 105733
QY 989 ccgaatagacctcaactatctbgacagctgtgcccgaacacccgtcctgtgtccagaccc 1048
|||||
Db 105734 CCGAATAGACCTCAACTATCTGCGAGTGTGGCGGACACCTGCGCTGTTCACACACCG 105793
QY 1049 cagcctgacctctatgtgcaactctgtgtcaaccactgtctttaaagcttgaactgtgag 1108
|||||
Db 105794 CAGCCCTGACCTCTATGGCAATCTGGGTCACCCACTGTCTTAAAGCTTGACTTCTGTGAG 105853
QY 1109 tttagacctgcccccccaaccacccctgcacatagtagctgtctcatgttgaactgtg 1168
|||||
Db 105854 TTTAGACCTGGCCCTCCACCCACCCCTGCACATATGAGCTATGCTCATGTGACTTGG 105913
QY 1169 aggcagatccagagacagctctccctcaactctgagaaaccttgacctctctgtatgaacac 1228
|||||
Db 105914 AGGCAGATCCAGGACACAGCTCCCTCACTTGGAGAACCTTGACTCTTGATGGAACAC 105973
QY 1229 agatggctgtctgggaaagaacttcaacctgaacttcaactgaagtcagactgaact 1288
|||||
Db 105974 AGATGGCTGCTGTGGGAAGAACTTTCACCTGACCTTGAGGTGACAGCTGCACT 106033
QY 1289 tcagaagaagtggaatttatatagctcatgttatttcaatggaacatgaagctctgtcga 1348
|||||
Db 106034 TCAGAAAGGTGGAATTTATATAGTCATTTATTTCAATGGAACATGAAGTTCTGCTGA 106093
QY 1349 gggctgagcagcagctgcatgtgaaataataataataaagtc 1393
|||||
Db 106094 GGGCTGAGCAGCAGCTGCAATGAAAAATATATATATCAATAAAGTC 106138

RESULT 6
AL354714.5
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome 1 c

Query Match 40.4%; Score 572.2; DB 2; Length 110000;
Best Local Similarity 98.6%; Pred. No. 2.9e-157;
Matches 577; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 869 cccaagccacactggagaggtgtgtctgtcggcgccgtgtctatcgaaacccagtgctatgtagt 868
|||||
Db 5554 CACATCTCATGCGCCAGGCTGCTGCTGCGGCCGCTCTATCAACCAACTGCTATGTATGT 5613
QY 869 ggcacacacagctgttgacgcccacacatgagagagaaagttatggccacagcatgt 928
|||||
Db 5614 GGCACACACAGCTGTGAGCGCCACCATGAGAGAGCAATTTATGGCCACGATGGT 5673
QY 929 ggtagacccctggggaacagctgtgcccgcctgtctctgaggggacagccctctgtc 988
|||||
Db 5674 GGTAGACCCCTGGGGAACAGTGTGGCGCGCTGTGAGGGGCGACAGGCGCTGTGCTTGC 5733
QY 989 ccgaatagacctcaactatctbgacagctgtgcccgaacacccgtcctgtgtccagaccc 1048
|||||
Db 5734 CCGAATAGACCTCAACTATCTGCGAGTGTGGCGGACACACCTGCGCTGTTCACACACCG 5793

QY 1049 cagcctgacctctatgtgcaactctgtgtcaccacactgtctttaaagcttgaactgtgtgag 1108
|||||
Db 5794 CAGCCCTGACCTCTATGGCAATCTGGGTCACCCACTGACTTAAAGCTTGACTTGTGAG 5853
QY 1109 tttagacctgcccccccaaccacccctgcacatagtagctgtctcatgttgaactgtg 1168
|||||
Db 5854 TTTAGACCTGGCCCTCCACCCACCCCTGCACATATGAGCTATGAGCTCATGTGACTTGG 5913
QY 1169 aggcagatccagagacagctctccctcaactctgagaaaccttgacctctctgtatgaacac 1228
|||||
Db 5914 AGGCAGATCCAGGACACAGCTCCCTCACTTGGAGAACCTTGACTCTTGTATGGAACAC 5973
QY 1229 agatggctgtctgggaaagaacttcaacctgaacttcaactgaagtcagactgaact 1288
|||||
Db 5974 AGATGGCTGCTGTGGGAAGAACTTTCACCTGACCTTGAGGTGACAGCTGCACT 6033
QY 1289 tcagaagaagtggaatttatatagctcatgttatttcaatggaacatgaagctctgtcga 1348
|||||
Db 6034 TCAGAAAGGTGGAATTTATATAGTCATTTATTTCAATGGAACATGAAGTTCTGCTGA 6093
QY 1349 gggctgagcagcagctgcatgtgaaataataataataaagtc 1393
|||||
Db 6094 GGGCTGAGCAGCAGCTGCAATGAAAAATATATATATCAATAAAGTC 6138

RESULT 7
AL590651/c
LOCUS
DEFINITION
Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
AL590651
AL590651.4 GI:13990236
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Harrison, E.
TITLE
JOURNAL
COMMENT
Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On May 7, 2001 this sequence version replaced gi:13751001.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
Project Information
Center project name: bA137A12
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: Plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 164729 bases at least Q40
Consensus quality: 165486 bases at least Q30
Insert size: 167063; sum-of-contigs
Insert size: 171878; 3.4% error; agarose-fp
Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality
coverage: 8.11x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9373: contig of 9373 bp in length
* 9374 9473: gap of 100 bp

QY	989	ccgataagactcaactatctcgcgaaggttcggcgacacttgctgtgttccagaccg	1048
Db	149516	CCGATAGACCTCAACTATTTCGCACAACTTCCGCCGACAACCTGGCTGTGTTCAGCACCG	149457
QY	1049	caggccctacctcatgtgcaatctcgggtcacccaactgtcttaagaactgtactctgttag	1108
Db	149456	CAGGCCTBACTCTATATGGCATTCTGGGTCACCACCTGTCTTAAGAATTGAATTCTGTGAG	149397
QY	1109	ttagaacctgcacctcccacccccacccctgcacatatgactagtctcatgtactgttg	1168
Db	149396	TTTAGACTTGCCCTCCACCCCCAACCCCTGGCATTATAGTAGTAGCATGTGACTTGG	149337
QY	1169	agcgcaagatccaagcacagcctccctcaactctggsgaaactgtactctcttaaygaaac	1228
Db	149336	AGGAGAGATTCACAGGCACAGCCCTCATCTTGGAGAACCTTGACTCTCTTATGGAACAC	149277
QY	1229	aagctggactgtcttgsgaaaagaactttcaactgagactcactgagtcacagtcagatt	1288
Db	149276	AGATGGCTGCTTGGGANAAGAACTTTTCAGCTTGACCTTCACTTGAGGTCAACTGCACTT	149217
QY	1289	tcgaaaggttgtaaatttatatagtcatttgtttaattcaatggaactyaaigtctgcga	1348
Db	149216	TCAGAAAGGTGCGAATTTTATATAGCATGTGTTATTCATGGAACATGAATTCGCTGCA	149157
QY	1349	gggctgagcagcactgycattgaaaaataataatacataaagtc	1393
Db	149156	GGGCTGAGCACACTGCGATTGAAAAATATATATCATATAAGTC	149112
RESULT	8		
LOCUS	AL591806	200822 bp	DNA linear PRI 30-JAN-2002
DEFINITION	Human DNA sequence from clone Rpl1-544M22 on chromosome 1, complete sequence.		
ACCESSION	AL591806		
VERSION	AL591806.16	GI:18476709	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 200822)		
JOURNAL	Harrison,E. Direct Submission Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
COMMENT	On Feb 1, 2002 this sequence version replaced g1:17902927. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one MJ3 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.: EMBL; SW.: SWISSPROT; Tr.: TREMBL; Wp.: WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chrl Rpl1-544M22 is from the library RPCL-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm		

Db	3608	-CAGCACCCGACACCGACTCTTATGCGAGTCGTGGGTATCCACACTCTCTTAATGCTTTTAA	3667
OY	1099	cttctgtyagttttagactctgccctcccacccccaacctgcacatafyagctatgctca	1158
Db	3668	CTTCTGTGTGTTTAAACACACAGACACAG------CAACTGGAGGATGATGCTAC	3715
OY	1159	tgttactctfygagcagagatccagcgacagctccctccactctfygagagacttga-cctctct	1217
Db	3716	TGTGACTTGTGAGCGCAGCAGACACAGCGTTCCCTTCACCTTGACAGACTTTAACTCTCT	3775
OY	1218	tgtatggaaacacagatbgygcctctfygagaaagaattccactgaagctcaactcyagctc	1277
Db	3776	TAAATGGAAACACAGCGCTTAATCAAGTCCATGGGAA--AGAAGAGACTTTGACCTGACCTC	3832
OY	1278	agacbcagcttccagaaagctfygaattccataagcaatgttltatccatbagaacttga	1337
Db	3833	AGATTTCACTTTCAGAAAGTGGGATTTTATATAGTCACTGTTTATTCATGAAACTGA	3892
OY	1338	agctctgctgagggctcgygcagcagcactgctgcatg----aaaataataatcatataagct	1393
Db	3893	AGTTATGCTGTAGGGCTGAGCGACACTGTGCATTGAAAAAATATATATATCATATATGTC	3952

RESULT	11			
AC084821				
LOCUS				
DEFINITION	AC084821	211771 bp	DNA	linear
ACCESSION	Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE.			
VERSION	AC084821			
KEYWORDS	AC084821.24	GI:18390260		
SOURCE	HTG; HTGS_PHASE2; HTGS_DRAFT.			
ORGANISM	house mouse.			
	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 211771)			
AUTHORS	Descamps,S., Gu,W. and Roe,B.A.			
TITLE	Mus musculus BAC Clone rp23-395h6			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 211771)			
AUTHORS	Descamps,S., Gu,W. and Roe,B.A.			
TITLE	Direct Submission			
JOURNAL	Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,			
	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,			
	OK 73019, USA			
COMMENT	On Jan 29, 2002 this sequence version replaced gi:16076995.			

```

----- Genome Center
Center: Department of Chemistry And Biochemistry
The University of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
  1 211771: contig of 211771 bp in length.
    Location/Qualifiers
      1..211771

```

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-395h6"
/clone_lib="PC1 Mouse BAC Library 23"
BASE COUNT      55622 a 48688 c 49207 g 58231 t      23 others
ORIGIN

```

Query match	23.18;	Score 326.8;	DB 2;	Length 211771;
-------------	--------	--------------	-------	----------------

Best Local Similarity 77.78; Pred. NO. 1.1e-84;
Matches 466; Conservative 0; Mismatches 112; Indels 22; Gaps 5;

QY	801	atlcagagcccaagcccaacttggagagtggtctctgggggcccgtgctatgaaaccaatgac	860
Db	2707	AGTCACGGCAGCTCTCATCTCTCAGGTGCTGTCTGGGGCCCGCCGCAATTGAATCTCAGTGC	2766
QY	861	tatgtatgtgcaagcagcaacagctgtgtgaagccacatagtgaagaagcaagttatgtgcaac	920
Db	2767	TATGTATATACACAGCGGAGGTGTGAGCCCAACATGAAACAAAGACAAGTTATGGCAT	2826
QY	921	agcaatgtgtgtagacccttggggaaacagtgtgtgtgtcccgctgtctctgtgagggccagcctc	980
Db	2827	AGCATGTGTGTTTACCCGTGGGGCACAGTGTGTGGCCCGCTGCTCCGAGGAGCAAGGCTTC	2886
QY	981	tgccttccccgaatagaacctcaactctctcgtgaagatttggcgcagcaactgccttggtctc	1040
Db	2887	TGCTTGTCTGAAATTGTATCTTCACATTCTTACACACAGATGCGCAACACCTGCTGTGTGTC	2946
QY	1041	cagcacccgacgagcctgaactctatgtgcaatctctgtgtcacaaccactgtcttlaagac--	1098
Db	2947	CAGCACCGCAGACCTGACCTCTATGGCAGTGTGGTGATCCACCTCTTAATAGCTTTTAA	3006
QY	1099	cttctctgtgagtttagacctgtgccctcccaaccccaacctgtgcactatgaagctagtgtctca	1158
Db	3007	CTTCTGTGTTGTTTAAACACACAGACACAGC-----CAACGTGGAGGTGATGTCTAC	3054
QY	1159	tgtaccttggagcagaagatcccaagcagcagctccctcaacttggagaaccttga--ctctct	1217
Db	3055	TGTACTTGTGAGCAGACACACAGCAGCGTTCCTTTCACTTGCAAGACTTTAACTCTCTCT	3114
QY	1218	tgaatgaacaacagatgtggtctgtcttggaaagaaccttacccttgagcttcaacctgtaggttc	1277
Db	3115	TAAATGGAACACAGGCTTAATCAAGTCCATGGGAA--AGAAAGACTTTGACCTGACCTC	3171
QY	1278	agacatcagtttcaagaagaagttgaaatttatatatatgtcatgtgttatcttcaatggaacctga	1337
Db	3172	AGATTTTCAGTTTACAGAAAAAGTGGGATTTTATATAGTCACTGTTTATTTTCATGAAAAACGA	3231
QY	1338	agttctctgaaggctgtgagcaagcctgtgattg----aaaataataatcaataaagct	1393
Db	3232	AGTTATCTTAGGGCTTGAGCAGCACCTGGCATTTGAAAAAAAATATATATATCATATATGTC	3291

RESULT	12
AC087229	213203 bp DNA linear HTG-05-FEB-2002
LOCUS	Mus musculus chromosome 1 clone rp23-191a19, WORKING DRAFT
DEFINITION	SEQUENCE, 4 unordered pieces.
ACCESSION	AC087229
VERSION	AC087229.15 GI:18497119
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Jiang,X., Song,L., Gu,W. and Roe,B.A.
TITLE	Mus musculus Chromosome 1 BAC Clone rp23-191a19
REFERENCE JOURNAL	Unpublished
AUTHORS	2 (bases 1 to 213203)
TITLE	Jiang,X., Song,L., Gu,W. and Roe,B.A.
REFERENCE JOURNAL	Direct Submission
AUTHORS	Submitted (19-DIC-2000) Department Of Chemistry And Biochemistry,

COMMENT On Feb 5, 2002 this sequence version replaced gi:18390237
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently


```

* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 11373: contig of 11373 bp in length
* 11374 11473: gap of unknown length
* 11474 43477: contig of 32004 bp in length
* 43478 43577: gap of unknown length
* 43578 126526: contig of 82949 bp in length
* 126527 126626: gap of unknown length
* 126627 213203: contig of 86577 bp in length.
FEATURES
    source
        1. 213203
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="fp23-191a19"
            /chromosome="1"
BASE COUNT  57858 a 50465 c 49150 g 55425 t      305 others
ORIGIN
Query Match      23.1%; Score 326.8; DB 2; Length 213203;
Best Local Similarity 77.7%; Pred. No. 1.1e-84;
Matches 466; Conservative 0; Mismatches 112; Indels 22; Gaps 5;

QY 801 attacagagccagccagctggagagtgctgctggagccgctgctatcagaaccagctgc 860
DB 50373 AGTCACAGCGACGCTCTCTCTCAAGCTCTGCTGGGCGCCGCCGATTCGATTCGATTCG
QY 861 tatgtatgagcagcagcagcagctgtgagcagcagcagcagcagcagcagcagcagcagc 920
DB 50433 TATGTATATACAGCAGCAGCAGCTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC
QY 921 agatgtgtgtatagaccctgtggagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 980
DB 50493 AGCATGGTGGTTAGCCGCTGGGACACAGTGTGGCCCGCTGCTCCGAGGACACAGGCGCTC
QY 981 tgccttgcgcgaatagaccctcaactatctgagcagctgtgagcagcagcagcagcagcagc 1040
DB 50553 TGCTTGTCTGCAATTCATCTCCACTTCTTACACAGATTCGCGCCACACCTGCGCTTTC
QY 1041 cagcagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1098
DB 50613 CAGCAGCGCAGACCTGACCTGTATGAGCAGCTGGGTGATCCACTCTCTTATATGCTTTAA
QY 1099 ctctctgtgaatttagacctgcccctccaccaccaccacctgccaactatgactagctctca 1158
DB 50673 CTCTCTGTGGTTAGACACAGACACAGC-----CAACTGGAGAGTGTGTCTAC 50720
QY 1159 tctgacttggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1217
DB 50721 TGTGACTTGTAGGACAGCAGCAGAGCGTTCCCTTACTTTCGAGAACTTTAACTTCTCT
QY 1218 tgaatgaacacagatgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1277
DB 50781 TATATGAAACACAGGCTTATATCAAGTTCATGAGAA---AGAAGAGACTTTGACCTGCAGCCTC 50837
QY 1278 agactcagcttccagaagaagtgaaatttatatagctatcattgttattcatgtgaactga 1337
DB 50838 AGATTTCAGATTTCAGAAAATTTTATATAGTACACTGTTTATTTTATTCATGAAAACATGA 50897
QY 1338 agcttcagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1393
DB 50898 AGTTATGCTGAGGCGCTGACGACGACCTGTCATGAAAAAAATATATATCATTAATATGTC 50957

```

```

RESULT 13
AC105589
LOCUS      AC105589      178418 bp      DNA      linear      HTG 09-JAN-2002

```

```

DEFINITION      Rattus norvegicus clone CH230-242B2. *** SEQUENCING IN PROGRESS
ACCESSION      AC105589
VERSION        AC105589.1 GI:18092812
KEYWORDS       HTG: HTGS_PHASE1.
SOURCE         Norway rat.
ORGANISM       Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE      1 (bases 1 to 178418)
AUTHORS        Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarella,J.,
Benton,D., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chaver,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Deen,A.L., Ding,Y., Dinh,H.H., Douthwalte,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Huliyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudh,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neel,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nokenko,S.,
Ogun,M., Okunolu,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Stinson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Wellington,S., Williams,G., Williamson,A., Wleczyk,R., Woden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 178418)
REFERENCE      Worley,K.C.
AUTHORS        Direct Submission
SUBMITTED      Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT        ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: NGI
Center clone name: CH230-242B2
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 132588 bases at least Q40
Consensus quality: 142233 bases at least Q30
Consensus quality: 152072 bases at least Q20
Estimated insert size: 141845; sum-of-contigs estimation

```

Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.8x in Q20 bases; sum-of-contrigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently consists of 65 contrigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contrigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 8302: contrig of 8302 bp in length
8402: gap of unknown length
8403 18472: contrig of 10070 bp in length
18473 18572: gap of unknown length
18573 24922: contrig of 6350 bp in length
24923 25023: gap of unknown length
25023 30990: contrig of 5877 bp in length
30990 38320: gap of unknown length
38320 38420: contrig of 7321 bp in length
38421 43244: gap of unknown length
43245 43344: gap of unknown length
43345 48228: contrig of 4884 bp in length
48229 48328: gap of unknown length
48329 54595: contrig of 6267 bp in length
54596 54695: gap of unknown length
54696 58551: contrig of 3856 bp in length
58552 58651: gap of unknown length
58652 62204: contrig of 3553 bp in length
62205 62304: gap of unknown length
62305 66587: contrig of 4283 bp in length
66588 70928: gap of unknown length
70929 71028: gap of unknown length
71029 75627: contrig of 4599 bp in length
75628 75727: gap of unknown length
75728 78372: contrig of 2545 bp in length
78373 78372: gap of unknown length
78373 80301: contrig of 1929 bp in length
80302 80401: gap of unknown length
80402 83675: contrig of 3274 bp in length
83676 83775: gap of unknown length
83776 87173: contrig of 3398 bp in length
87174 87273: gap of unknown length
87274 90196: contrig of 2923 bp in length
90197 90296: gap of unknown length
90297 92065: contrig of 1769 bp in length
92066 92165: gap of unknown length
92166 94612: contrig of 2447 bp in length
94613 94712: gap of unknown length
94713 98755: contrig of 4043 bp in length
98756 98855: gap of unknown length
98856 102543: contrig of 3688 bp in length
102544 102643: gap of unknown length
102644 105833: contrig of 3190 bp in length
105834 105933: gap of unknown length
105934 108499: contrig of 2566 bp in length
108500 108599: gap of unknown length
108600 111672: contrig of 3073 bp in length
111673 114662: contrig of 2890 bp in length
114663 114762: gap of unknown length
114763 116145: contrig of 1383 bp in length
116146 116245: gap of unknown length
116246 118147: contrig of 1902 bp in length
118148 118247: gap of unknown length
118249 120579: contrig of 2332 bp in length
120580 120679: gap of unknown length
120680 122618: contrig of 1939 bp in length
122619 122718: gap of unknown length

122719 124539: contrig of 1821 bp in length
124540 124639: gap of unknown length
124640 126651: contrig of 2012 bp in length
126652 126751: gap of unknown length
126752 128460: contrig of 1709 bp in length
128461 128560: gap of unknown length
128561 130592: contrig of 2032 bp in length
130593 130692: gap of unknown length
130693 132864: contrig of 2172 bp in length
132865 132964: gap of unknown length
132965 134881: contrig of 1917 bp in length
134882 134981: gap of unknown length
134982 136682: contrig of 1701 bp in length
136683 136782: gap of unknown length
136783 139138: contrig of 2356 bp in length
139139 139238: gap of unknown length
139239 140757: contrig of 1519 bp in length
140758 140857: gap of unknown length
140858 142558: contrig of 1701 bp in length
142559 142658: gap of unknown length
142659 144303: contrig of 1645 bp in length
144304 144403: gap of unknown length
144404 144490: contrig of 1087 bp in length
144491 145491: gap of unknown length
145492 147171: contrig of 1381 bp in length
147172 147271: gap of unknown length
147273 148639: contrig of 1368 bp in length
148640 148739: gap of unknown length
148740 150295: contrig of 1556 bp in length
150296 150395: gap of unknown length
150396 151036: contrig of 1110 bp in length
151037 151505: gap of unknown length
151506 151605: gap of unknown length
151606 152813: contrig of 1208 bp in length
152814 152913: gap of unknown length
152914 155423: contrig of 2510 bp in length
155424 155523: gap of unknown length
155524 156880: contrig of 1357 bp in length
156881 156980: gap of unknown length
156981 158045: contrig of 1065 bp in length
158046 158145: gap of unknown length
158146 159255: contrig of 1110 bp in length
159256 159355: gap of unknown length
159356 160417: contrig of 1062 bp in length
160418 160517: gap of unknown length
160518 161644: contrig of 1127 bp in length
161645 161744: gap of unknown length
161745 163180: contrig of 1436 bp in length
163181 163280: gap of unknown length
163281 164483: contrig of 1203 bp in length
164484 164583: gap of unknown length
164584 166039: contrig of 1456 bp in length
166040 167290: gap of unknown length
167291 167290: contrig of 1151 bp in length

Query Match 21.8%; Score 308.4; DB 2; Length 178418;
Best Local Similarity 74.3%; Pred. No. 3e-79;
Matches 447; Conservative 0; Mismatches 141; Indels 14; Gaps 4;

QY 789 gcttttgatcattacagccagccagctggaggtgtctcgagccctgtctatc 848
DB 41294 GCTGATGTTCTGAGTACGACGCTCTCACTCAGTCTCTCTGCGGCGCCGCAATT 41333
QY 849 gaaccacagtcattatgtagtgcagcagcagcagctgtgagcgcacacatagaagagca 908
DB 41354 GAGTTCACAGTCTAGCTATGATACAGACAGAGTGTGAGCGCACACATAGACAGAGAGA 41413
QY 909 agttatgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 968
DB 41414 ACTTACGCGCACAGCATGCTGTGACCCCTGGCGCACAGTGTGCGCCAGCTGTCCGAG 41473
QY 969 gggcagcagccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 1028
DB 41474 GGACGAGCGCTTGCCTTGCCTGCAATGTATTCACACTTCTGCAACAGATGCGCAACAC 41533

QY 1029 ctgcctgtctccagaccgagcctgagcctctatggaacttggtgacccactgctc 1088
|||||
Db 41534 TTGCCGTGTTCAGCATGACAGACCTGACCTTATGGCAGTCT---CTGGCAGCTCTCC 41590
QY 1089 taagacttgactctgtgagctttagaacttgcctccaccacccaccctgacactagag 1148
|||||
Db 41591 TATGTTTAAATTTCTGTGGTTGA-----GACAGCAAAACACACCGCACTGTGAG 41643
QY 1149 ctggtctcatgtgacttgagagcagatccagacagcactccctcacttgtagaacct 1208
|||||
Db 41644 GGTATGCTACCGGAGCTGTGAGGCTGACCCAGCAGCGGTTCCTTCACCTTCAGAGG-CT 41702
QY 1209 tgactctctgagagacagatgagcctgctgctgaggaagaacattccactgagcttc 1268
|||||
Db 41703 TAACTCTCTTAATGACACAGGCTCAATCACTGAGGAA---AGAGAGACCTTTGA 41759
QY 1269 cctgaagtcagactgagcttccagaaagctggaatttlatatagctatgcttcat 1328
|||||
Db 41760 CCGATCTCAGATTTGAGTTTCAGAAAATTTTATATGTCACCTTTTATTTCAAT 41819
QY 1329 ggaactgagctctgctgagggctgagcagcactgagcactgagaaataataatcacta 1388
|||||
Db 41820 GAAAACTGAAGTGTGCTGAGGCTGAGCAGCAGCTGCAATGAAAAAATATATATATC 41879
QY 1389 aa 1390
Db 41880 AA 41881

RESULT 14

LOCUS G13436 441 bp DNA linear STS 04-JUN-1996
DEFINITION human STS WI-12338, sequence tagged site.
ACCESSION G13436
VERSION G13436.1 GI:1127545
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the Unigene collection.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STS
JOURNAL Unpublished

COMMENT

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: TATATTTTCAATGCCAGTGTG
Primer B: AACCTTCACCTGAGCTTCACC
STS size: 130
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:

Protocol:
Template: 10 ng
Primer: each 5 pm
dNTPs: each 4 mM
Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

FEATURES
Derived from dbEST (genbank accession T78621).
Location/Qualifiers

source

1..441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="750.8 CR from top of Chr1 linkage group"

STS

primer_bind

primer_bind

BASE COUNT 107 a 95 c 130 g 103 t 6 others
ORIGIN

Query Match

Best Local Similarity 98.4%; Score 280; DB 11; Length 441;
Matches 313; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 1079 ccactgtcttaagacttga-cctcgtgagctttagaactgcccctccacccaccct 1137
|||||
Db 317 CCCACTGCTTAACTGACCTTCTGTGAGTTAGACCTGCCCTCCACCCCACTTCC 258

QY 1138 gccac-ctatgagctagctgctcatgctgacttgagagcagatccagacagctccctca 1196
|||||
Db 257 GCCANCTATGAGCTAGTGCATGTGACTTGGAGGCGAGNCCAGGACGACGCTCCCTCA 198

QY 1197 ctggagaaacttgactctctttagtgaacacagatggctgcttgggaaagaacttc 1256
|||||
Db 197 CTTGGAGAACCTTGACTCTCTTATGGAACACAGATGGCTGTGGGAAAGAACTTTC 138

QY 1257 acctgagcttcacactgagctgagcagacttccagaaagctggaatttlatatagctat 1316
|||||
Db 137 ACCTGAGCTTCACCTGAGCTGACAGCTGACACTTTCAGAAAGTGAATTTATATAGCAT 78

QY 1317 tgttatttcaatgaactgaacttctgctgagggctgagcagcagcactggaataat 1376
|||||
Db 77 TGTATTATTCATGGAACCTGAGTTCGTGGA-GGCTGAGCAGCAGCTGCATGGAATAAT 19

QY 1377 ataataatcctaagctca 1394
|||||

Db 18 ATATATATCATTAAGTCA 1

RESULT 15

LOCUS G72919 847 bp DNA linear STS 08-AUG-2001
DEFINITION MARC 2849-2850:991933517.1 SCF - porcine spleen Sus scrofa STS
genomic, sequence tagged site.
ACCESSION G72919
VERSION G72919.1 GI:15146949
KEYWORDS STS.

SOURCE

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 847)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L. and
Keeler, J.W.

REFERENCE

AUTHORS Single nucleotide polymorphism (SNP) discovery in expressed porcine
genes
TITLE Unpublished (2001)
JOURNAL

COMMENT

Contact: Freking BA
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4278
Fax: 402 762 4173
Email: freking@email.marc.usda.gov

THIS PAGE BLANK (USPTO)

XX 20-JUL-1998: 98US-0093350.
PR
XX
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
XX
PI Croce CM;
DR
DR WPI; 2000-171195/15.
DR P-PsDB; AAY68739.
XX
XX Novel nitrlilase homologs used as diagnostic and therapeutic reagents
PT for the detection and treatment of cancer -
XX
PS Claim 6; Fig 6; 25pp; English.

CC The present sequence represents the coding region of human, murine,
CC Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC human and mouse NIT1 genes are members of an uncharacterised
CC mammalian gene family with homology to bacterial and plant nitrilases.
CC The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC for fusion proteins in which the Fhl domain is fused with a Nit domain.
CC In mouse and humans, FHIT and NIT are encoded by two different genes,
CC localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC chromosomal fragile site FRAB, is often altered in most common forms
CC of human cancer. The Nit1 protein overcomes the mutated inactivation
CC of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC and analogues of them, and antibodies are used as diagnostic and
CC therapeutic reagents for the detection and treatment of cancers.

Query Match	99.7%	Score 1411.8	DB 21	Length 1416
Best Local Similarity	99.98%	Pred. No. 0		
Matches 1414, Conservative	0	Mismatches	2	Indels 0, Gaps 0

[illegible]

QY	541	gtcacggtgcgcgtgaaacagcaaggggcagctgaatggccacttacaggaagacaactcgt	600
Db	541	gtcacggtgcgcgtgaaacagcaaggggcagctgaatggccacttacaggaagacaactcgt	600
QY	601	gtgacgctagaagatctccagggcaggggcctatagtgtgaaagcaacttaccatgctgggc	660
Db	601	gtgacgctagaagatctccagggcaggggcctatagtgtgaaagcaacttaccatgctgggc	660
QY	661	ccgaatcttgaggtcaacctgcttaagcacaccacagcagaagtgtgtctaagtgtctgctatg	720
Db	661	ccgaatcttgaggtcaacctgcttaagcacaccacagcagaagtgtgtctaagtgtctgctatg	720
QY	721	acacgtgcgttcccgcaaacctctctctgtgcattgtgcctcaaacctgagcagagatactact	780
Db	721	acacgtgcgttcccgcaaacctctctctgtgcattgtgcctcaaacctgagcagagatactact	780
QY	781	atccctcagaacttttgaatccatattacagccacagccaccatgtgagtggtgtgtcgtcgagcc	840
Db	781	atccctcagaacttttgaatccatattacagccacagccaccatgtgagtggtgtgtcgtcgagcc	840
QY	841	gtgtatccgaaaccccaatgttatgtatgtgcacagacagatggtgtgacgccaacataaga	900
Db	841	gtgtatccgaaaccccaatgttatgtatgtgcacagacagatggtgtgacgccaacataaga	900
QY	901	agagagacaggttatgagccacagacatggtgtgaagccctctggggagacaagtgtgtccgcgt	960
Db	901	agagagacaggttatgagccacagacatggtgtgaagccctctggggagacaagtgtgtccgcgt	960
QY	961	gctcttgaggggcagagggcctctgtccctgtgcccgaatgaacctcaactatctgtcgacagttgc	1020
Db	961	gctcttgaggggcagagggcctctgtccctgtgcccgaatgaacctcaactatctgtcgacagttgc	1020
QY	1021	ggccacacccctgacctgtgttccacagaccgcgcgtcgtacctatagtgcacatctcgggttcacc	1080
Db	1021	ggccacacccctgacctgtgttccacagaccgcgcgtcgtacctatagtgcacatctcgggttcacc	1080
QY	1081	caatgtctttaaagacttgaacttctgtgaatttagaactgcctcccaaccccccaacctgcgc	1144
Db	1081	caatgtctttaaagacttgaacttctgtgaatttagaactgcctcccaaccccccaacctgcgc	1144
QY	1141	actctgagcctagtcctcatatgtgaacttggaagcagagatccacagacagcctccctcaactg	1200
Db	1141	actctgagcctagtcctcatatgtgaacttggaagcagagatccacagacagcctccctcaactg	1200
QY	1201	gagaacacttgaactctctctgtgaacacagatgggctgtcttgggaaagaaacttcaacct	1266
Db	1201	gagaacacttgaactctctctgtgaacacagatgggctgtcttgggaaagaaacttcaacct	1266
QY	1261	gagcttcaaccgtgaggtcagactgtcagatttcaagaaggtgggaatttataatgaatcgtct	1322
Db	1261	gagcttcaaccgtgaggtcagactgtcagatttcaagaaggtgggaatttataatgaatcgtct	1322
QY	1321	tattctatgnaaaacttgaagtctctgtctgaggggcgtgagcagcaactgtgattgaaataata	1386
Db	1321	tattctatgnaaaacttgaagtctctgtctgaggggcgtgagcagcaactgtgattgaaataata	1386
QY	1381	taatcatataaagtcataaaaaaaaaaaaaaaaaaaaaa	1416
Db	1381	taatcatataaagtcataaaaaaaaaaaaaaaaaaaaaa	1416

AAFI6257	2	AAFI6257	standard; cDNA; 1382 BP.
XX		AAFI6257;	
AC		13-MAR-2001 (first entry)	
XX			
DT			
XX			
DE			
XX			
XX			
KW			
RW			
neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;			

KW vulnery; gastrointestinal; nephrotropic; antinefctive; gynaecological,
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200055174-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05988.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCT INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-587513/55.
 DR P-PSDB; AAB57054.
 XX
 PT prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 PS Claim 1; Page 1124; 2338pp; English.
 XX
 CC AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,
 CC nephrotropic, antinefctive, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive, .
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.
 XX
 SE Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

Query Match	93.3%;	Score 1321;	DB 21;	Length 1382;
Best Local Similarity	97.3%;	Pred. No. 0;		
Matches 1376;	Conservative	0;	Mismatches 6;	Indels 32;
				Gaps 22;

QY 1 gccacatcgctgcggcctmtctgtctccaaacgcgcctccgagatcgaccctgcgaatg 60
Db 1 gccacatcgctgcggcgccttctgtctccaaacgcgcctccgagatcgaccctgcgaatg 60
QY 61 ttctggcctatctctcaatgtatgaaacctaaccctatccgctggcggcgtctggctta 120
Db 61 ttctggcctatctctcaat-----gcctggcctta 89
QY 121 tcacacagcgcctcctcaacagattcctctgcctctgtgtctctgagactccgatatacc 180
Db 90 tcacacagcgcctcctcaacagattcctctgcctctgtgtctctgagactccgatatacc 149
QY 181 tctcaagtaacttctgtgtctcagcgcgcacagagcagatgctatctctctctctgcg 240
Db 150 tctcaagtaacttctgtgtcctcagcgcgcacagagcagatgctatctctctctctgcg 209
QY 241 aactgcgcctctgtgcgtgtgtgcacaggtaacatctgacgcagacagaacagaactta 300
Db 210 aactgcgcctctgtgcgtgtgtgcacaggtaacatctgacgcagacagaacagaactta 269
QY 301 aaacatgtcgtcagctcgtgttcagaaaggctgcacagactggtgtcctgcgtgtcttcctcg 360

Dh	270	aaacatgtctgagctgtgtcttcgagagagctgcagagcttgggtgtcctgcgtcttctctgc	325
Qy	361	ctgagagcatttggacttcatctgtcacaagggaccctgtcagagacgcttacaacctgtctgaacac	420
Dh	330	ctgagagcatttggacttcatctgtcacaagggaccctgtcagagacgcttacaacctgtctgaacac	389
Qy	421	tgggttgggaaaccttttggaaagaatacaccagcttgcaggggaatgtgtgactctgtctgt	480
Dh	390	tgggttgggaaaccttttggaaagaatacaccagcttgcaggggaatgtgtgactctgtctgt	449
Qy	481	ctttgggtgggtttcccatgagcctgtgcagacacccaggggaagagacttcaagaatactacaatt	540
Dh	450	ctttgggtgggtttcccatgagcctgtgcagacacccaggggaagagacttcaagaatactacaatt	509
Qy	541	gtcacgtgtctctcgtcaaaacagcaaaaggggcagttatgtgtccacttacaaggaacacactgtt	600
Dh	510	gtcacgtgtctctcgtcaaaacagcaaaaggggcagttatgtgtccacttacaaggaacacactgtt	569
Qy	601	gtgacgtgtgagatattccagggccaggggacctatgtgtgaaagacaactctacaatgtcctggc	660
Dh	570	gtgacgtgtgagatattccagggccaggggacctatgtgtgaaagacaactctacaatgtcctggc	628
Qy	661	ccagttcttggatccaacctgtcaagcacaccacccagccaggaagtgtgtctagtctgtgtcatgt	720
Dh	629	ccagttcttggatccaacctgtcaagcacaccacccagccaggaagtgtgtctagtctgtgtcatgt	688
Qy	721	acatgtcgtttccctcgaaactctctctcgtcagtttgtctaaagcttgagacagatatactact	780
Dh	689	acatgtcgtttccctcgaaactctctctcgtcagtttgtctaaagcttgagacagatatactact	748
Qy	781	atccttcaagcttttggatcccatlacaagcccagcccaacttggagagtgtgtcctgcggcc	840
Dh	749	atccttcaagcttttggatcccatlacaagcccagcccaacttggagagtgtgtcctgcggcc	808
Qy	841	gtgtcatctcgaaacccagttgtcatgtgtgtgcagcacagcaaggtgtgacgcacacatgaga	900
Dh	809	gtgtcatctcgaaacccagttgtcatgtgtgtgcagcacagcaaggtgtgacgcacacatgaga	868
Qy	901	agaagagcaagattatgagccacagcatagtgtgtaagacccttggggaaacaggtgtgagccct	960
Dh	869	agaagagcaagattatgagccacagcatagtgtgtaagacccttggggaaacaggtgtgagccct	928
Qy	961	gctctgaggggcccagggccctctgtcgttgcgccgaataagaccctacaactatctgtgcagactgtc	1020
Dh	929	gctctgaggggcccagggccctctgtcgttgcgccgaataagaccctacaactatctgtgcagactgtc	988
Qy	1021	gccgaacacctgcgtgttctccagcaccccgagccgctgaactctatgtgcgaatctgtgtcaac	1080
Dh	989	gccgaacacctgcgtgttctccagcaccccgagccgctgaactctatgtgcgaatctgtgtcaac	1048
Qy	1081	cactgtcttgaagacttgaactctgttgcaggtttgaacctgtgccttccaccccccaacctgtcc	1140
Dh	1049	cactgtcttgaagacttgaactctgttgcaggtttgaacctgtgccttccaccccccaacctgtcc	1108
Qy	1141	actatgagctagtgtcctcatgtgaacttggagagcagatattcaagccacagcttccccctactgt	1200
Dh	1109	actatgagctagtgtcctcatgtgaacttggagagcagatattcaagccacagcttccccctactgt	1168
Qy	1201	gagaaacctgaactctcttgaatgagaacagatgggctgtctgttggaaagaacttcaacct	1260
Dh	1169	gagaaacctgaactctctcttgaatgagaacagatgggctgtctgttggaaagaacttcaacct	1228
Qy	1261	gagcttcaacctgaggtlcaagactggaagttltaagaagaagtgtgaatttataatgttaatttt	1320
Dh	1229	gagcttcaacctgaggtlcaagactggaagttltaagaagaagtgtgaatttataatgttaatttt	1288
Qy	1321	tatttcatgagaacttgaagtctgtctgtgagggcctggaacacatctgtgcatgtgaaaaatafat	1380
Dh	1289	tatttcatgagaacttgaagtctgtctgtgagggcctggaacacatctgtgcatgtgaaaaatafat	1348
Qy	1381	tattcatataagtcataaaaaaataaaaaaataaaaaa 1414	
Dh	1349	tattcatataaaaaaataaaaaaataaaaaaataaaaaa 1382	

Db	839	tatatagcattggttatttcacatggaactgaaagttctctgagagcctgagcagacttg	898
Oy	1366	catgaaaaataataatcatcaaaagtc	1393
Db	899	catgaaaaataataatcataaagtc	926
RESULT 4			
ID	AAD25458	standard; cDNA; 1214 BP.	
XX			
AC	AAD25458;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Xenopus laevis Nlt1 cDNA.		
XX			
KW	Frog; Nlt1; cytosolic; neuroprotective; cellular pathway; therapy;		
XX	apoptosis; proliferative disorder; degenerative disease; ss.		
OS	Xenopus laevis.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	180..1046	
FT		/*tag= a	
FT		/product= "Frog Nlt1 protein"	
XX			
PN	WO200187958-A2.		
PD			
XX	22-NOV-2001.		
XX			
PE	15-MAY-2001; 2001WO-US15664.		
XX			
PR	16-MAY-2000; 2000US-204713P.		
XX			
PA	(UYJE-) UNIV JEFFERSON THOMAS.		
XX			
PL	Croce C, Brenner C, Pekarski Y;		
XX			
DR	WPI; 2002-082984/11.		
DR	P-PSDB; AAE15792.		
XX			
PT	Isolated cDNA encoding human, mouse, frog and yeast Nlt2 proteins,		
PT	useful to find molecules that mimic or antagonize Phit interaction for		
PT	the treatment of proliferative or degenerative diseases		
XX			
PS	Claim 21; Page 60; 61pp; English.		
XX			
CC	The invention relates to isolated nucleic acids comprising a fully		
CC	defined cDNA nucleotide sequence encoding human, Xenopus laevis and		
CC	mouse Nlt2 proteins. Nlt and Phit proteins are encoded as fusion		
CC	proteins in vertebrates and as separate polypeptides in vertebrates.		
CC	Nlt and Phit interact physically and functions in same cellular pathways.		
CC	Molecules which bind Nlt2 and mimic or antagonise Phit interaction are		
CC	used to treat diseases in which activity of Nlt2 protein is altered in		
CC	a mammal. Phit mimics induce apoptosis and are particularly useful to		
CC	treat proliferative disorders, whilst Phit antagonists promote cell		
CC	proliferation and are particularly useful to treat degenerative disease.		
CC	The present sequence is frog Nlt1 cDNA.		
XX			
SO	Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;		
Query Match 25.2%; Score 356.2; DB 24; Length 1214;			
Best Local Similarity 64.5%; Pred No. 2, 3e-88;			
Matches 532; Conservative 0; Mismatches 293; Indels 0; Gaps 0			
Oy	245	gccacctgctgctgtgtgccaggtaacatcgacgcagacaagaacaagaattaaac	304
Db	197	gcccttgatgcgtgtgctgcagatgacttcaacctctctgtaagagagaattgcac	256
Oy	305	atgttctgagctggtttcgaagctgcgcagaactgggtgcctgctgcttctcgtctga	364

Dd	257	gtgttcgcgcgtcatccggagagcctcgcgggcgcgtccgccttcgaagtggtttcttcgcgga	316
Oy	365	ggaattcgtacttcattcgaaggagaccctcgcagagacgcgtacacctcgtctcgaaccactcggg	424
Dd	317	agacctgacatacatcgggggcagcaattcggagagacgcctgactcgtctcgtcgtacttaaa	376
Oy	425	tcggaaaccttttggaaatacaaccacgctctcgcagggaaatgttgactctcgtcctctt	484
Dd	377	tcggagacaaccaattcagtcgttacaccaccaactcgcagggagtgctgggtcctcgtcttccct	436
Oy	485	gggtggtttccatctagcgttcggccaagacgttggagcagactcagaaatacatcaattcga	544
Dd	437	gggggggattctatcagaaagagaccacactcgggcacagcagccaaacgcattccattccaa	496
Oy	545	cgctctcgtcgaacagcaaaaggggcagtgatggtgcaccttcaagaaagacacatctgttga	604
Dd	497	ctgtgttggtggaacaacaacagggcacaatagatactggtgataccgcaaggtccacactgttga	556
Oy	605	cgtagagaattccagggcagggccctatgctgtgaagaacaaactctaccaatcgtcggccag	664
Dd	557	cgtagaacttggaagatgagtgcttaactcagagagagcaqattccacacctcccgagacaga	616
Oy	665	tcctgagatcaacctcgtcagcacacacagaggaagaattgttctagcctgctctgctatgaact	724
Dd	617	gcttattccgcacacacacacacacacacagagaaagatcttgcctcgggggtgtgttaacgacct	676
Oy	725	gcggtctccctgaacctctctctcgtgcattcgtgcataagcttggagcagagagatacttaccatcc	784
Dd	677	cgctctccacgaattctcctctgtcctcgtcgcacaaagagacagaaactcttcaattacc	736
Oy	785	ttcagcttttggatccattcaacagccacagccacactcggaggtgtgtgcctcggggcccgctgc	844
Dd	737	ttctgcctctccacctcactactcactcgtctcgtgcacattgggaggtgtgtctagaagcccgctgc	796
Oy	845	tatcgaaaccccaagcgtcatatgattggtgagagagacagaggtgttgcgcgaccacatgagaaagag	904
Dd	797	catagaaaccccaagcgtcatatgattggtgagagagacagagacagacacacacatgagaaagag	856
Oy	905	agcaagttatctgccaacagcactgtgtgtgtagacccctcgtggagaaacaggtgtgtgcgcgcgtgc	964
Dd	857	gaacttccatctatgtcatcgtctatgtgtgtgtagaccctgtgggtcgtgtcatattggccaatgcga	916
Oy	965	tcgaggggccaagcctctgcctcttgcgcgaatagacactcaactatctcgcagacagttgcgcgcg	1024
Dd	917	ggaaggaacagaaatcatgtatgctgtgagatcttgacattccctacatctgagcgtgtgtagcgcg	976
Oy	1025	acacctgctcgtgtgttcacagacacgcagagcgtcgtgcctcttatgtgaa	1069
Dd	977	ggaactgcccgtgtgtggaagcagccgcaagactcgtcatctglatggaa	1021
RESULT 5			
ABLI12225			
ID	ABLI12225 standard; cDNA; 1495 BP.		
XX	ABLI12225;		
AC	ABLI12225;		
XX	26-MAR-2002 (first entry)		
DT	26-MAR-2002 (first entry)		
XX	Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.		
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.		
XX	Drosophila; developmental biology; cell signalling; insecticide;		
KM	Drosophila; developmental biology; cell signalling; insecticide;		
KN	pharmaceutical; gene; ss.		
XX	Drosophila melanogaster.		
OS	Drosophila melanogaster.		
XX	WO200171042-A2.		
PN	WO200171042-A2.		
XX	27-SEP-2001.		
PD	27-SEP-2001.		
XX	23-MAR-2001; 2001WO-US09231.		
PF	23-MAR-2001; 2001WO-US09231.		
XX	23-MAR-2000; 2000US-191637P.		
PR	23-MAR-2000; 2000US-191637P.		

QY 253 tggctgtgcccaggtgaatcagacgcccagacaagcaacttaaaactgtgctg 312
 DB 1200 tggctgtgcccaggtgaatcagacgcccagacaagcaacttaaaactgtgctg 1259
 QY 313 agctgtgctgagagctgcccagactgagctggtgctgctgctgctgctgagcattg 372
 DB 1260 agctgtgctgagagctgcccagactgagctggtgctgctgctgctgagcattg 1319
 QY 373 acttcaatgacagcagccctgacagagcgtacacactgtctgaacacactggtggtggaac 432
 DB 1320 acttctgtggtcagagagcgcgacccaactatttgactctcccgaggtgtgagcgcgagct 1379
 QY 433 ttcttgaaagaatacaccacagctgtccagaggaatgtgactctgtctgtctgtgtgtgt 452
 DB 1380 taatgtgcagatgacccggaattgtgcgaatgtcacacaatttggattccctgtgtgtgcg 1439
 QY 493 tccatgagcgtgtgcccagactgtggtgagcagactcagaagaatcttaacttgcacgtgtgc 552
 DB 1440 tgcacgagcggaaacga-----tcaaaaaatcttcaacgctcactgttttgc 1484
 QY 553 tgaacagcaagggcagctagtggtccactacagagaagacacactctgtgtgacgtgaga 612
 DB 1485 tcaacagagaagggagactagcagcagatatacagaagaagctgcacatgctgtgtataga 1544
 QY 613 ttccagggcaggggctatgtgtgaaagcaacctacatgctgtggtccaggtcctgagct 672
 DB 1545 cttaaa---aggtctgctcactacgcaatcagatatacagcttaccggtgatactgtctgagc 1601
 QY 673 cactctcagcacacacagcagaagcaatgtgtctagctgtctgtctatgacgtgtcc 732
 DB 1602 gcccaatgtgagcactccagtttgcagatatagggcttcaagatttgcacgactcgttttg 1661
 QY 733 ctgaactctctcgtgactgtgtcctaaagctgtgagcagagataacttaacttccctcagct 792
 DB 1662 ctgagcgcggcgtgtgtctcagagaagctgtgtgcgaattgtttaacataccatccgcacat 1721
 QY 793 ttggtatccattacagccagccagcactgtggaagtggtgtgtcgtggtggtccgtcatgaa 852
 DB 1722 tcaacatcagcaacggttaagcgcgactgtggaatctctcgtcgtggtggtccaggtcagataga 1781
 QY 853 ccagatgcatgtgtgtgacagcagcagcagctgtgtgagccagccacttggaaagagcgaagt 912
 DB 1782 ctcaatgcttctgtgtcgtcgtggtcagataggttggcacaacagaaagcagcagaagt 1841
 QY 913 atggccacagctgtgtgtgtagaacctcctgtggaagcaagtggtgc---ccgtgtctgagag 969
 DB 1842 ggggccaacagcatgtatcgttagccctgtggaagaaagtaactgtgactgtgagcagagc 1901
 QY 970 ggcgaagctctgtcctgtccgaataagactacatactgtggaagctgtggtgcgacacac 1029
 DB 1902 agcttataatagcagcagcaggaagtgagacttccgtgtccttaactctcttatacagaac 1961
 QY 1030 tgcctgtgtccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1067
 DB 1962 tgcctgtgtccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1999

RESULT 7
 ABL20222
 ID ABL20222 standard; DNA; 5692 BP.
 XX
 AC ABL20222;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical; gene; ds.
 OS Drosophila melanogaster.
 XX

PN MO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX
 XX Venter JC, Adams M, Li PMD, Myers EM;
 PI
 XX
 XX WPI; 2001-658660/75.
 DR
 XX
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX
 PS Claim 1; SEQ ID NO 12139; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins
 CC (AB157737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 XX Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;
 SQ

Query Match 12.9%; Score 182.8; DB 23; Length 5692;
 Best Local Similarity 54.4%; Pred. No. 5.3e-40;
 Matches 445; Conservative 0; Mismatches 352; Indels 21; Gaps 3;

QY 253 tggctgtgcccaggtgaatcagacgcccagacaagcaacttaaaactgtgctg 312
 DB 2859 tggctgtgcccaggtgaatcagacgcccagacaagcaacttaaaactgtgctg 2918
 QY 313 agctgtgctgagagctgcccagactgagctggtgctgctgctgctgagcattg 372
 DB 2919 agctgtgctgagagctgcccagactgagctggtgctgctgctgagcattg 2978
 QY 373 acttcaatgacagcagccctgacagagcgtacacactgtctgaacacactgtggtggtggaac 432
 DB 2979 acttctgtggtcagagagcgcgacccaactatgagctctccagaggtgtgagcgcgagt 3038
 QY 433 ttcttgaaagaatacaccacacttgcagaggaatgtgactcgtgtctcctgtgtgtgt 492
 DB 3039 taatgtgcagctatccgggaattgtgcgaatgtgcacaacaatttggattccctgtgtgtgcg 3098
 QY 493 tccatgagcgtgtgcccagactgtggtgagcagactcagaagaatctcaaatgtcaactgtcgc 552
 DB 3099 tgcacgagcggaaacga-----tcaaaaaatcttcaacgctcactgttttgc 3143
 QY 553 tgaacagcaagggcagctagtggtccacttaccagagaagaacactctgtgtgagcagtagaga 612
 DB 3144 tcaacagagaaggggagactagtcagcagctatatacagaagaagctgcacatgtttgagttaaga 3203
 QY 613 ttccagggcaggggctatgtgtgaaagcaacttaccatgtggtggtggtccagctgtgagt 672
 DB 3204 cttaag---aggtctgctcactacggaatacagatatacagtaagtcggtggtatgactgtgagc 3260
 QY 673 cactctgacacacacagcagcagaagatgtgtcagctgtcgtctatgacatgcgtgtcc 732
 DB 3261 gcccaagtgagcactccagctgtgcccagatagagcttcaagatttgcacagactcgttttg 3320
 QY 733 ctgaactctctcgtgactgtgtcccaagctgtgagcagagataacttaacttccctcagctt 792

```
Db 3321 ctagacccggtgctgctcaggaagctggtcccaattgttaacataccatccgcat 3380
Oy 793 ttgagtcacattacagccaccacactggaggtgtgtgtggtggccgtgtcatcga 852
Db 3381 tcaatacagcaaccggttaagcgcacttggaaatcctctcggtggccagacataga 3440
Oy 853 cccagtgctatgtatgtagcagacagtgltgagccacatgagaagaagcaagt 912
Db 3441 cccaatgtctgtgtcgtcgtcagatagttgtgacacacaggaagcgacagagt 3500
Oy 913 atggccacagcatgtgtgtgagacccctgtggagacagtggtgc---ccgtgtctgag 969
Db 3501 ggggcccacagcatgctgtagccctgtgggaacgtactgtgactgacgagcagag 3560
Oy 970 ggcacagcctcgtcctgtccgaatagacctcaactatgctgcagacagttgcgcgcac 1029
Db 3561 agcttgatataagcagcgcgaggtggaaccttcgtcgttcaatctcgtatcagaca 3620
Oy 1030 tgcctgtgtccagacccgagcctgacctatgtgc 1067
Db 3621 tgcctgtcgtcgacatcgtcgaaacgacatctacgcc 3658
```

RESULT 8

AAD25457
ID AAD25457 standard; cDNA; 1292 BP.

AC AAD25457;

DT 26-MAR-2002 (first entry)

DE Mouse Nlt2 cDNA.

KM Mouse; Nlt2; cytosolic; neuroprotective; cellular pathway; therapy;
apoptosis; proliferative disorder; degenerative disease; ss.

OS Mus sp.

Key Location/Qualifiers
FH 46.876
FT /*tag= a
CD5 /product= "Mouse Nlt2 protein"

PN WO200187958-A2.

PD 22-NOV-2001.

PF 15-MAY-2001; 2001WO-US15664.

PR 16-MAY-2000; 2000US-204713P.

PA (UYJE-) UNIV JEFFERSON THOMAS.

PI Croce C, Brenner C, Pekarski Y;

DR WPI; 2002-082984/11.

DR P-PSDB; AAE15791.

PT Isolated cDNA encoding human, mouse, frog and yeast Nlt2 proteins,
useful to find molecules that mimic or antagonize Flt1 interaction for
the treatment of proliferative or degenerative diseases

PS Claim 9; Page 59-60; 61pp; English.

CC The invention relates to isolated nucleic acids comprising a fully
defined cDNA nucleotide sequence encoding human, Xenopus laevis and
mouse Nlt2 proteins. Nlt and Flt1 proteins are encoded as fusion
proteins in invertebrates and as separate polypeptides in vertebrates.
Nlt and Flt1 interact physically and functions in same cellular pathways.
Molecules which bind Nlt2 and mimic or antagonise Flt1 interaction are
used to treat diseases in which activity of Nlt2 protein is altered in
a mammal. Flt1 mimics induce apoptosis and are particularly useful to

CC treat proliferative disorders, whilst Flt1 antagonists promote cell
proliferation and are particularly useful to treat degenerative disease.
CC The present sequence is mouse Nlt2 cDNA.
XX Sequence 1292 BP; 354 A; 292 C; 289 G; 357 T; 0 other;

Query Match 8.0%; Score 113.8; DB 24; Length 1292;
Best Local Similarity 51.7%; Pred. NO. 3.4e-21;
Matches 284; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

```
Oy 516 gacgagactcagaaatctacaattgtccgtgctgtgaaacagagggcagtagtg 575
Db 307 gagagctcggggaactgtataatccctgtgtgtgtgtgtgtgtgtgtgtgtgt 366
Oy 576 gccacttaaggaagacacatctgtgtgacgtagaagattccagggcagggcctatgt 635
Db 367 gtaagacaggaagaaatccatctgtttgacatgtatgttccctgggaaattcgttcaa 426
Oy 636 gaaagcaactcaccatgtcctgtggccagcttctgagtcacgttcagcacaccagcagc 695
Db 427 gatctaaacattgagccctgtgtatagttctccacatttgat---agccttactgc 483
Oy 696 aagattgtctagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 755
Db 484 aagatggcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 543
Oy 756 caagctggagcagagatacttaccttccctcagcttggatggatccattacacagcagc 815
Db 544 caaagagctcagcagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 603
Oy 816 cactggagagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 875
Db 604 cactggagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 663
Oy 876 gacagtgltgagcgcacacatgagaagagcagatgtatgtgtgtgtgtgtgtgtgt 935
Db 664 tctcctgtcgtgagacaaagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 723
Oy 936 cccttgggagacagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 995
Db 724 ccttgggggagagcttccaaacaaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 783
Oy 996 gactcaactatctgcagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1055
Db 784 gactgaagaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 843
Oy 1056 gacctctat 1064
Db 844 gacctctat 852
```

RESULT 9

AAF84214
ID AAF84214 standard; cDNA; 1001 BP.

AC AAF84214;

DT 12-JUN-2001 (first entry)

DE Human hNlt3-ase coding sequence.

KM Human; hNlt3-ase; Cushings adrenal gland tumour;

KM nitric hydrolytic enzyme; nitrlase; ss.

OS Homo sapiens.

Key Location/Qualifiers
FH 99..929
FT CDS /*tag= a
FT /product= "Human hNlt3-ase"

PN CN1277998-A.

```

XX 27-DEC-2000.
PD AAD25456
XX 30-MAY-2000; 2000CN-0116221.
XX 30-MAY-2000; 2000CN-0116221.
XX 30-MAY-2000; 2000CN-0116221.
XX (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.
XX
XX Xu X, Qian B, Zhang X;
XX WPI; 2001-245678/26.
DR P-PSDB; AAB80984.
XX
XX Human nitrite hydrolytic enzyme protein and its coding sequence -
XX
XX Claim 1; Page 16-17; 20pp; Chinese.
XX
XX The present invention relates to human hnit3-ase (nitrlase) protein,
XX which is expressed in Cushings adrenal gland tumour, and its coding
XX sequence (AA84214 and AAB80984). The present invention also relates to a
XX preparation method of said protein and nucleic acid sequence, and a
XX method of detecting human hnit3-ase nucleic acid sequence and polypeptide
XX in sample.
XX
XX Sequence 1001 BP; 264 A; 235 C; 245 G; 257 T; 0 other;
XX

```

```

Query Match 7.2%; Score 102; DB 22; Length 1001;
Best Local Similarity 50.4%; Pred. No. 5.6e-18;
Matches 277; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```

```

QY 516 gagcagactcagaagaatctcaattgtcaactgtcgtcgtgacagaaaggcgagtagt 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 360 gagagatctgggaattataacacactgtgctgtgttggtgacctgtagaacttacta 419
QY 576 gccactcaggaagacacactctgtgtgacgtagagattccaggcgagggccatgtgt 635
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 420 gcaaatatagaagaacacactctgtgtgacatgtgttccctggaataataacttca 479
QY 636 gaaagcaactctaacactgcctggtgcccagcttgaatcactgtcgaacacacagcaggc 695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 480 gaactcaaacatgtgctcgggtgtagattt---ctccacattgtactccttactgc 536
QY 696 aagatgtctagctgtctgtactatgacatgcggttccctgaactctctgtgcatlgtgt 755
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 537 agagtgtgtctgtgacatctgtacacatgcggttgcagagcttgacacaaatctacgca 596
QY 756 caagcttgagcagagatctactactatccttcaagcttlttgatcattacagccagcc 815
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 597 cagagaagctgcagcagctgtgtgtatataccagagacttlaactgaccacacgacacgc 856
QY 816 cactggagaggtgtgtctgcggccgtgtactcgaacccagctgtatgtgtgacagca 875
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 657 catggagagtactactcagcagaagccgggtgtgtatataccagcgtgtgtgtgacagcc 935
QY 876 gcacagtgtgacgcccacatgagaaagagacagattatgcccacagatgtgtgtgac 935
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 717 tctctctgccggagatcaaacgctctctatgttgcctgtggagacacagacccgtgtgac 935
QY 936 cccctggggaacagtggtgtgcccgtgtctctgtgaggggcccagccctgtcctgtccgata 995
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 777 ccttgggggagaggttctcagcacaagctgtgacagagaagaacaaatcgtgtatcagaacta 836
QY 996 gacctcaactatctgacacagctgtgacccagacactgtcgtgtccagcagcgagcct 1055
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 837 gacctgaagaagcgtgctgtaataatcgcacagcaaatcccgctttagacagaaagcatal 896
QY 1056 gacctctatg 1065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 897 gacctctatg 906

```

```

RESULT 10
AAD25456
ID AAD25456 standard; cDNA; 1359 BP.
XX
XX AAD25456;
AC
XX 26-MAR-2002 (first entry)
DT
XX
XX Human Nit2 cDNA.
DE
XX
XX Human; Nit2; cytosolic; neuroprotective; cellular pathway; therapy;
XX apoptosis; proliferative disorder; degenerative disease; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 22..852
XX FT /tag= a
XX FT /product= "Human Nit2 protein"
XX
XX W0200187958-A2.
XX
XX 22-NOV-2001.
XX
XX 15-MAY-2001; 2001WC-US15664.
XX
XX 16-MAY-2000; 2000US-204713P.
XX
XX (UYJE-) UNIV JEFFERSON THOMAS.
XX
XX Croce C, Brenner C, Pekarski Y;
XX
XX WPI; 2002-082984/11.
XX
XX P-PSDB; AAE15790.
XX
XX Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,
XX useful to find molecules that mimic or antagonize Phit interaction for
XX the treatment of proliferative or degenerative diseases
XX
XX Claim 3; Page 59; 61pp; English.
XX
XX The invention relates to isolated nucleic acids comprising a fully
XX defined cDNA nucleotide sequence encoding human, Xenopus laevis and
XX mouse Nit2 proteins. Nit and Phit proteins are encoded as fusion
XX proteins in invertebrates and as separate polypeptides in vertebrates.
XX Nit and Phit interact physically and functions in same cellular pathways.
XX Molecules which bind Nit2 and mimic or antagonise Phit interaction are
XX used to treat diseases in which activity of Nit2 protein is altered in
XX a mammal. Phit mimics induce apoptosis and are particularly useful to
XX treat proliferative disorders, whilst Phit antagonists promote cell
XX proliferation and are particularly useful to treat degenerative disease.
XX
XX The present sequence is human Nit2 cDNA.
XX
XX Sequence 1359 BP; 388 A; 284 C; 312 G; 374 T; 1 other;
XX

```

```

Query Match 7.2%; Score 102; DB 24; Length 1359;
Best Local Similarity 50.4%; Pred. No. 6.5e-18;
Matches 277; Conservative 0; Mismatches 270; Indels 3; Gaps 1;
QY 516 gagcagactcagaagaatctcaattgtcagctgtcgtcgtgacagaaaggcgagtagt 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 283 gagagatctgggaattataataacacactgtgtgtgtgtgtgacctgtagaacttacta 342
QY 576 gccactcaggaagacacacactgtgtgacgttagagattccaggcgagggccatgtgt 635
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 343 gcaaatatagaagaatcactctgttgcacattgtatgttccctggaaaaaataactttaa 402
QY 636 gaaagcaactctaacatgtcgtggcccagctgtgtgacacactgtcagcacacagcagc 695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 403 gaactcaaacatgtgctcgggtgtagattt---ctccacattgtactccttactgc 459
QY 696 aagatgtctagctgtctgtatgacatgtcggttccctggaactctctctgtgattgct 755

```

Db 460 agagtggtctggtgcatctgctacgacatcggtttgcagagcttgacaaatctacgca 519
 Oy 756 caagctggagcagagatactactactctcctcagcttttgatccaatagccagcc 815
 Db 520 cagaagagctgcccagctgtgtatatacagagagctttaaatacaccatggacagcc 579
 Oy 816 cactggagagtggtgtgctggcgagccgtgctatcgaaaccagctgcatatgagcagca 875
 Db 580 catggagagttacttcacgagaaagccggtgtgtgataacagtgatgtggtgacagcc 639
 Oy 876 gcaacatgtgagcagccacatgagaaagagcaagatattgcccacagcagcatgtgtgagac 935
 Db 540 tctccctcccgagatgacaaagccctcctatgtgtcctggggagacacagcagcgtgtgagac 699
 Oy 936 ccttgaggagacagtggtgtgcccgtctcctgagggcgagccgtctgctgtcccgcaata 995
 Db 700 ccttgaggaggaggttctcagcagcaagctgtgacagaaagcaatcgtgtatctagacata 739
 Oy 996 gacctcaactatctgagacagtgctgcgcagacacctgctgtgttcagcagccagcagcc 1055
 Db 760 gacctgaagaagctgtgctgtaataacgacagcaaatcccgctttttagacagaaagcatca 819
 Oy 1056 gacctctatg 1065
 Db 820 gacctctatg 829

RESULT 11

AAH07371 standard; cDNA: 505 BP.

AAH07371:

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:4206.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602

full-length cDNAs defined in the specification, and for the detection

and/or diagnosis of the abnormality of the proteins encoded by the

full-length cDNAs -

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the 5'-end
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

Sequence 505 BP; 105 A; 120 C; 167 G; 108 T; 5 other;

Query Match 7.1%; Score 101.2; DB 22; Length 505;
 Best Local Similarity 96.3%; Pred. No. 6,7e-18;
 Matches 103; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 4 cactgcgtgcgagccttctgctccagacgagccctcgagatcgagaccctgcatgtt 63
 Db 3 ctctgcgtgcgagccttctgctccagacgagccctcgagatcgagaccctgcatgtt 62

Oy 64 tggctatctctcaatgtagaacctactccctcgtgcgcgag 110
 Db 63 tggctatctctcaatgtagaacctactccctcgtgcgcgag 109

RESULT 12

ABL13787 standard; cDNA: 898 BP.

ABL13787:

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polynucleotide SEQ ID NO 35843.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ss.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PMD, Myers EW;

WPI; 2001-656860/75.

New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signalling and cell-cell

interactions -

Claim 1; SEQ ID NO 35843; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB157737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 898 BP; 243 A; 236 C; 231 G; 188 T; 0 other;

Query Match 7.1%; Score 101.2; DB 23; Length 898;
 Best Local Similarity 50.6%; Pred. No. 8.8e-18;
 Matches 272; Conservative 0; Mismatches 263; Indels 3; Gaps 1;

QY 531 atctacaatgtcagctgctgcgaacgaagggagtagtgccacttacagaag 590
 DB 344 atctacaacacccctgcagctgttgtaaccactggtgactcttggtccaagcatcgcaag 403
 QY 591 acacactgtgtgacgtgagatgccagggcagggcctatgtgtgaagcaactctacc 650
 DB 404 atgcactctcttgacataagatgtcgaagg--tgacattcgcttcaagaggtcggaagc 460
 QY 651 atgcccgggcccagctcttgagctcactgttcagacacacgaagcaagattgtctagct 710
 DB 461 ctgtccgacgggcaatgattccaccatcatcaagtagatggtccacaagattgtgcatcgcg 520
 QY 711 gtctgtatgacatgctgctccctgaactctctctgcatctgctcaagctgaagtcgaagcag 770
 DB 521 atctgtacgatatcttcgactcgaagagatggtcggaagctctatcgcaagcagctcgagc 580
 QY 771 atacttaccatcttcacgcttcttgatccattacacagccacccactgtggagtgctg 830
 DB 581 atgatactatctccgctgctgattccaacatgacacactgttcactgtcactgtggagctatg 640
 QY 831 ctgctgggcccgtctatcgaaaccagtgctatgtagtgcagacagacagctgtgacgc 890
 DB 641 cagcgatcccggtgcgaatgcaacacactcttctgtgtccaaatcaccagcccggtgac 700
 QY 891 caacatgagaagagagaattatgcccacagatgtgtgtagacccttggggaaacagtg 950
 DB 701 acaagcgccgagatattagcctatggtccattccatggtgtgatatccatggtgccaagtg 760
 QY 951 gtggccgctgtctctgaggggccaagcctctgtccttgccgaatagacactcaatctcgt 1010
 DB 761 cagcagagtgccagtgaaagcgaggaatcgtgtgtggtcgcatatagattcttcgagtg 820
 QY 1011 cgaagattgcccgaacacactgctgtgttccacagcagcagcgtcactctatgca 1068
 DB 821 gagcaggtgctcagcagattcccgctcttctgtgcaaaagcgtctagatctgtacgca 878

RESULT 13

AB13786
 ID AB13786 standard; cDNA; 2953 BP.

XX AB13786;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35840.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR MPI; 2001-656860/75.

XX P-PSDB; ABB69683.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PS interactions -

XX Claim 1; SEQ ID NO 35840; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB157737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2953 BP; 692 A; 688 C; 776 G; 797 T; 0 other;

Query Match 7.1%; Score 101.2; DB 23; Length 2953;
 Best Local Similarity 50.6%; Pred. No. 1.6e-17;
 Matches 272; Conservative 0; Mismatches 263; Indels 3; Gaps 1;

QY 531 atctacaatgtcagctgctgcgaacgaagggagtagtgccacttacagaag 590
 DB 1399 atctacaacacccctgcagctgttgtaaccactggtgactcttggtccaagcatcgcaag 1458
 QY 591 acacactgtgtgacgtgagatgccagggcagggcctatgtgtgaagcaactctacc 650
 DB 1459 atgcactctcttgacataagatgtcgaagg--tgacattcgcttcaagaggtcggaagc 1515
 QY 651 atgcttggcccagctcttgatcactgttcagcacacacgaagcaagatgtgtctagct 710
 DB 1516 ctgtccgacggagcaatgattccaccatcatcaagctagatgtgcacaagattgcatcgcg 1575
 QY 711 gtctgtatgacatgctgctccctgaactctctctgcatctgctcaagctgtgagcagag 770
 DB 1576 atctgtcagatattctcagatgagagatgtgcaggtctatctgcaacgagcgttcgag 1635
 QY 771 atacttaccatcttcagcttcttgatccattacagccacgcccacttgggaagtgctg 830
 DB 1636 atgatactatctccgctgcatcttcaacatgacacactgtgcactgtcgtggagctatg 1655
 QY 831 ctgctggcccgtctatcgaaaccagtgctatgtagtgcagacagacagtggtgacgc 890
 DB 1696 cagcgatcccggtgcgaatgcaacacactcttctgtgtcaacaatcacacagcccggtgac 1755
 QY 891 caacatgagaagagagaattatgcccacagatgtgtgtagacccttggggaaacagtg 950
 DB 1756 acaagcgccgagatattagcctatggtccatcccatctgtgtgaatcattggtcgcaagtg 1815
 QY 951 gtggccgctgtctctgaggggccaagcctctgcttccggaatagacactcaatctcgt 1010
 DB 1816 cagcagagtgccagtgaaagcgaggaatcgtgtgtggtcgcatatagattcttcgagtg 1875
 QY 1011 cgaagattgcccgaacacactgtgtgttccacagcagcagcgtcactctatgca 1068
 DB 1876 gagcaggtgctcagcagattcccgctcttctgtgcaaaagcgtctagatctgtacgca 1933

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249221.
PR 17-NOV-2000; 2000US-0249222.
PR 17-NOV-2000; 2000US-0249223.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249225.
PR 17-NOV-2000; 2000US-0249226.
PR 17-NOV-2000; 2000US-0249227.
PR 17-NOV-2000; 2000US-0249228.
PR 17-NOV-2000; 2000US-0249229.
PR 17-NOV-2000; 2000US-0249230.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-483426/52.
P-PSDB; AAM88455.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -
PS
PS Claim 1; SEQ ID NO 6296; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;

Query Match 6.7%; Score 95.2; DB 22; Length 539;
Best Local Similarity 95.9%; Pred. No. 3.2e-16;
Matches 94; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Query 109 ggcctggcctacacacgcctcctcacacattccttgcctctgtgtcctgactcc 168
|||||

Db 371 ggctgggctcatcaccaggcctccacagattccctgcccctctctgtgckggnctcc 430
Qy 169 ggataacctcaactctcaactcttctgtgctcagcccaag 206
Db 431 ggataacctcaactctcaactcttctgtgctcagcccrpg 468

Search completed: April 28, 2002, 04:52:22
Job time: 3740 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2002, 01:30:45 ; Search time 1522.25 seconds
(without alignments)
12554.895 Million cell updates/sec

Title: US-09-357-675C-1
Perfect score: 1416
Sequence: 1 gccaccatcgcgcgcgcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	939	66.3	992	9	AL520767
C 2	927.6	65.5	960	9	AL522373
C 3	888.6	62.8	1374	11	AK002269
4	886.6	62.6	2993	11	AK004988
5	845	59.7	890	9	AL520768
6	844	59.6	844	9	AL529152
C 7	795.8	56.2	853	9	AL562604
C 8	772.4	54.5	830	9	AL580840
C 9	759.6	53.6	916	9	AL563355
C 10	758	53.5	846	10	AL563355
11	758	53.5	861	9	AL559163
12	748.6	52.9	928	10	AL559163
13	724.4	51.2	793	10	AL559163
14	704.4	49.7	787	10	AL559163
C 15	699.2	49.4	764	10	AL559163
16	686.8	48.5	746	10	AL559163
17					AL559163

18	678.4	47.9	813	10	BT755493
C 19	666.2	47.0	736	9	AT668782
20	665.2	47.0	946	10	BT757823
21	654.6	46.2	791	10	BT776560
22	643.4	45.4	781	10	BT532265
C 23	638	45.1	710	9	AT1797380
C 24	625	44.1	625	10	BT111300
C 25	619.4	43.7	828	10	BE895409
C 26	616.4	43.5	690	9	AA909728
27	609.2	43.0	875	9	AL547581
28	603.8	42.6	900	10	BF968272
C 29	598.2	42.2	778	10	BF968495
C 30	589.2	41.6	635	10	BF448631
C 31	584	41.2	586	9	AT1797259
C 32	583.6	41.2	900	10	BT432389
C 33	579.8	40.9	612	9	AT1831417
C 34	577.2	40.8	657	10	BT1916460
35	572.8	40.5	827	10	BF696074
36	572	40.4	579	10	BT111712
37	557.4	39.4	576	9	AW956706
C 38	557	39.3	559	10	BT411822
C 39	554.2	39.1	579	9	AT159188
C 40	552.8	39.0	560	9	AW182514
C 41	548.2	38.7	869	10	BT6963013
C 42	548	38.7	550	10	BT411736
C 43	547.4	38.7	550	9	AT017543
C 44	544.8	38.5	872	10	BT647584
45	535.6	37.8	10	BE392424	
					BT755493 603027349
					AT668782 wC14B12.x
					BT757823 603030404
					BT776560 602663693
					BT532265 602561355
					AT1797380 wE87d12.x
					BT111300 iD96f12.x
					BE895409 601434778
					AA909728 oM47g11.s
					AL547581 AL547581
					BF968272 602269658
					BF968495 6022771268
					BF448631 7n90d11.x
					AT1797259 wE86e02.x
					BT432389 602495419
					AT1831417 wI64b12.x
					BT1916460 603182527
					BF696074 601852067
					BT111712 iD96f12.y
					AW956706 EST368776
					BT411822 iF26h04.x
					AT159188 tQ42g05.x
					AW182514 xJ18b03.x
					BT6963013 602828020
					BT411736 iF25d10.x
					AT017543 oU35c05.x
					BT647584 603378164
					BE392424 601308982

ALIGNMENTS

RESULT 1	AL520767	992 bp	mrna	linear	EST 13-FEB-2001
LOCUS	AL520767	LRT1_NFL004_NBC2	Homo sapiens	cdna	clone CS0DB002YD05 3
DEFINITION	AL520767	prime, mRNA sequence.			
ACCESSION	AL520767	AL520767.1	GI:12784260		
VERSION	AL520767				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
COMMENT	Unpublished (2001)				
CONTACT	Genoscope				
Genoscope	Centre National de Sequencage				
BP 191	91006 EVRY cedex - France				
Email:	segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
Location/Qualifiers					
1. 992					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone_lib="CS0DB002YD05"					
/sex="male"					
/tissue_type="neuroblastoma cells"					
/lab_host="DH10B"					
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA					
was primed with a NotI-oligo(dT) primer. Five prime end					
enriched, double-stranded cDNA was digested with Not I and					
cloned into the Not I and Eco RV sites of the pCMVSPORT 6					
vector. Library was normalized. Library was constructed					
by Life Technologies. Contact : Feng liang life					
Technologies, a division of Invitrogen 9800 Medical Center					
Drive Rockville, Maryland 20850, USA fax : (1) 301 610					
8371 Email : filiang@life.com url :					
http://fulllength.invitrogen.com"					

BASE COUNT

229 a 267 c 268 g 222 t 6 others

ORIGIN

Query Match 66.3%: Score 939; DB 9; Length 992;
Best Local Similarity 98.7%: Pred. No. 1.8e-148;
Matches 962; Conservative 5; Mismatches 5; Indels 3; Gaps 2;

386 ggaacctgagagagacgtacacgtctgtaaacacgtgggtggaacatttggaaacata 445
|||||
984 GGACCTCGAGAGACGCTACCTGCTCTACACACCTGGGTGGAAC--TTTTCRAGRATA 927
|||||
446 caaccaagcttgcagggaatgtggaactgtgctgtccttgggtgtgttccatgagctg 505
|||||
926 CACCCAGCTTGGCCAGGGAATGTGAGACTGCGCTCTTGGGGGCTTCCATGAGCGGTG 867
|||||
506 ccaagaactggagagagacagaaaattacaaatgttgcagtgctgtgtaaacagcaag 565
|||||
366 CCAAGACTGGAGACACTCAGAAAATCTACAAATGTTCAGCTGCTCAACGCAAAAG 807
|||||
566 ggcagtagtggcactacaggaagacacatctgtgtgacgtatgagattccagggcag 625
|||||
806 GGCAGTAGTGGCCACTTACAGGAAGACATCTGTGTGACGTAGATTCAGAGGCAAGG 747
|||||
626 gctcatgtgtgaaagcaacttaccatgcttgggcccagcttctgagtcacctgtcagca 685
|||||
746 GCTATGTGTGAAGCAACTTACCAATGCTGGGCCAGTCTTGAGTCACTGTCAGCAC 687
|||||
686 accagagagcaagatgtgtctagctgtctgtcatgacatgctgtccctgaactctct 745
|||||
686 ACCAGAGGCAAGATGTGTGTAGCTGTGTGATGACATGGGTCCCTGAACCTCTCTCT 627
|||||
746 ggcattgtgtcaagctgtggcagagatacttaccttcccttcaagcttggatccattac 805
|||||
626 GGCATTGTGCTCAAGCTGGAGCGAGATACTTACTTACTTCAAGCTTTTGGATCCATTAC 567
|||||
606 agggccagagcccaactgtggaggtgtgtgtgctgggcccgtgtctatgaaacccagtgctatgt 865
|||||
566 AGGCCAGGCCCACTGAGAGGTGTGCTGCGGGGCCCTGCTATCGAACCACGAGCTATGT 507
|||||
866 agtggca -gcagacagatgttgagagccacatgagagagagagagagatgttgccacaaga 924
|||||
506 AGGCCAGGCCAGACAGATGTGTGAGCCACCATGAGAGAGAGCAATTTATGGCCACAGCA 447
|||||
925 tgggtgtgaacccttggaggaacagtggtgtgcccgtgtctctgagggccagggcctctgc 984
|||||
446 TGGTGTAGACCCCTGCGGGAACAGTGTGTGGCCGCTGCTGTAGGGGCGCAGGCTCTGCC 387
|||||
985 ttgcccgaatagaactcaactatctgtgagcaagttgcgcggaacacctgtgtgttccagc 1044
|||||
386 TTGCCGGAATAGACCTCACTATCTGCGACAGTTGGCCGACACCTGCTGTGTTCCAGC 327
|||||
1045 accgagagctgaactctatgtgcaatctgtgtcaaccacatgtgtcttaagaactgagcttgc 1104
|||||
326 ACCGCGAGGCTGACTCTATGGCAATCTGGGTACCCACTGTTTAAGACTTGACTTTCG 267
|||||
1105 tgaatttaagactgtccctcccaaccccaacccctgtgccaactatgagctagtgctatgtac 1164
|||||
266 TGAAGTTTAAACCTGCCCCCTCCACCCGCCACCTGGCAGCATATAGAGTGCATATGTGAC 207
|||||
1165 ttgagagagagatcagagacagctccctcaacttggagaaacctgtgactctcttgaatga 1224
|||||
206 TTGGAGGCGAGATCCAGGACAGCTCCCTCACTTGGAGAAACCTTGACTCTCTTGTATGA 147
|||||
1225 aaacagatgtgctgtctgtggaagaagaacttcaactgaacttcaactgtgtgtcaagctgc 1284
|||||
146 ACACACATATGGCTTTGGGAAAAGAACTTTCACCTGAGCTCTCMCTGAGGTAGACTTC 87
|||||
1285 agttcagaagaagtggaatttatatagtcattgtttatcttcaatggaaactggaactgcg 1344
|||||
86 AGTTTCAGAAAGGTGAATTTTATATAGTCAATTGTTTATTTCATGGAACCTGAAGTTCTG 27
|||||
1345 ctgagagctgagcag 1359
|||||

DB 26 CTGAGCGCTGAGCAG 12

RESULT 2
AL522373/C 960 bp mRNA linear EST 13-FEB-2001
LOCUS AL522373 LTI_NFL004_NBC2 Homo sapiens cDNA clone CSDB008YM02 3
DEFINITION prime, mRNA sequence.
ACCESSION AL522373
VERSION AL522373.1 GI:12785866
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homini; Homidae; Homo.
1 (bases 1 to 960)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDB008YM02"
/clone_11b="LTI_NFL004_NBC2"
/sex="male"
/tissue="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com"

BASE COUNT 225 a 259 c 255 g 217 t 4 others

ORIGIN

Query Match 65.5%: Score 927.6; DB 9; Length 960;
Best Local Similarity 98.9%: Pred. No. 1.5e-146;
Matches 951; Conservative 4; Mismatches 4; Indels 3; Gaps 2;

398 gacgtacacacctgtgtgaacacactgggttggaacatttggaaatatacacccagctgc 457
:|||||
960 RACGCTACACCTGTCTGAACACACTGGGTG--AAATTTTGAAGATACACCAGCTTGC 903
|||||
458 cagggaatgtgacttgcgtgcttgccttgggtgttccatgaagcgttgccaaagactggga 517
|||||
902 CAGGGAATGTGACACTGTGCTGCTGGGTGGTTCATGAGCGTGGCCAGACTGGGA 843
|||||
518 gcaagctcagaagaatcacaatgttcaagtgctgtcgtgaacagcaagggcagtgatgtgc 577
|||||
842 GCAGACTCAGAAAATCTACAAATGTTCACGTGCTGCGAACACCAAGGGCGAGTGTGTC 783
|||||
578 caactaagaagaagacacatctgtgtgagctagagattccaaggcagggccctatgtga 637
|||||
782 CACTTACAGGAAGACACATCTGTGTGACATGAGATTCAGGGCAAGGCGCTATGTGTA 723
|||||
638 aagcaacttacatagctgtggccagcttctgagtcacctgttcaagcacccagcagga 697
|||||
722 AAGCAACTTACATACCTCGGGGCCAGTCTTGAGTACTGTGCAGACACGAGGCA 663
|||||
698 gatgtctagctgtctgctatgacatgacgtgtccctgaactctctgtgcatgtgctca 757
|||||
662 GATTGCTTACGCTGCTCTATGACATGC -GTTCCCTGAAGCTCTCTGCGATTTGCTCA 604
|||||

OY	758	agctggagcagagataactatccatctccttcagccttttgatccatctacagggccagccca	817
Db	603	AGCTGGACACAGAGATCTTACTTACCTTACCTTTCAGCTTTTGGATCCATTACAGGCCACGCCCA	544
OY	818	ctgggaagtgctgtctgctgggcccgtgctatcgaaaccacagtgctatgtatgtgcagcagc	877
Db	543	CTGGGAGGTTGCTGGCGGGCCCGTGATATGAAACCCAGTGCTCTTGTATGTGGCAGCAGC	484
OY	878	acagtgtgagcgcacacatgaaagaagacaagttatgtgcacagcatgtgtgtagacc	937
Db	483	ACAGTGTGGACGCCACCATGAGAGAGACAGTATATGGCCACACATGATGTTGTAGACC	424
OY	938	ctgggaagacagtggtggccgctgtctcgtgaggggcgagacctctgcttggccgaatga	997
Db	423	CTGGGGAACAGTGGTGGCCCGCTGCTGTGAGGGGSCAGGCCCTTCCTTTGGCCGATATGA	364
OY	998	cctcaacatctctgacacagttgacgcagacaactgctgtgttcacagaccgagagctga	1057
Db	363	CCTCAACATATCTGCACAGTTTGGCGCAGACCTGCTGTGTTCCAGCACCGCAGGCTTGA	304
OY	1058	cctctatgagcaatctgggtgcacacacatgltcttaagacttgaactctctgtgagttagacct	1117
Db	303	CCTCTATGGCATATCTGGGTGCACCCACTGCTCTTAAACATGTGAATCTTGATGTGTAGACT	244
OY	1118	ggccctcccaacccccacacctgacactatgagtgatgtctcatgtgaacttggagggcaagat	1177
Db	243	GCCTCTCCACCCCCACCCCTGCACATATGAGCTAGTGTCAATGATCACTTGGAGGAGAT	184
OY	1178	ccaggcagacgtccccctcaacttggagaacacttgaactctcttgaatgaacaagatggct	1237
Db	183	CCAGGCACAGCTCCCTCTCMTCTTGGAGAACCTTGAATCTCTTGATGAAACACAGATGGCT	124
OY	1238	gcttgggaagaagaacttccatccatgagcttccacactgaggtcagacgtcagttccagaagg	1297
Db	123	GCTTGGAAGAACTTTCACCTGAGCTTACACTGAGGTCAACACGCAATTCACAAAGG	64
OY	1298	tggaaattatatagtcatgtgttatttcatatgaaactggaagttctgtgagggctgagc	1357
Db	63	TGGAAATTTAATATAGTCATTTGTTATTTTCATGGAACATGAACTTCGTGAGAGGTTGAGC	4
OY	1358	ag 1359	
Db	3	Ab 2	
RESULT	3		
	AK002269		
LOCUS	AK002269	1374 bp	mRNA linear HTC 19-JAN-2002
DEFINITION	Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610007B07:nitrlase 1, full insert sequence.		
ACCESSION	AK002269		
VERSION	AK002269.1	GI:12832130	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus.musculus (Strain:C57BL/6J) adult male kidney cDNA to mRNA, clone:0610007B07.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (sites)		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2 (sites)		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		

PUBMED REFERENCE	11042159
TITLE	3 (sites)
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,D., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kasaiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
JOURNAL	RIKEN Integrated sequence analysis (RISA) system-384-format
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	11076861
AUTHORS	4 (sites) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5 (bases 1 to 1374)
AUTHORS	Aadachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arawaka,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hangaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koza,S., Kutlbera,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numataki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schirml,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yamanae,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-Jun-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic sciences center and genome science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGACGGCCGCCAACATCGATCTTTTATTATTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAAGAGAGACGATCAAGAGCTCAATTAATTAATTAACCCTCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES	Location/Qualifiers
SOURCE	1. 1374 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="MGI:1891931" /db_xref="taxon:10090" /cclone="0610007B07" /sex="male" /tissue_type="kidney" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult"
gene	1. 1374 /gene="Ntlt1"
misc_feature	1. 1374 /note="data source:MGI, source key:MGI:1350916,"

MEDLINE	20530913
PUBMED	11076861
REFERENCE	4 (sites)
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5 (bases 1 to 2993)
AUTHORS	Adachi,T., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiroaka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kakusawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Schirral,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunushi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGACGGCCGCCCACTCGATGTATTTTATTATTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGACGATCCAGACACTCATTTATTTAATTAAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES	Location/Qualifiers 1..2993 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="MGI:1895066" /db_xref="taxon:10090" /clone="1300013F05" /sex="male" /tissue.type="liver" /clone.lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 169..1041 /gene="Niti1" 169..1041 /gene="Niti1" /note="data source:MGI, source key:MGI:1350916, evidence:ISS nitrilase 1 putative" /codon.start=1 /protein_id="BAB3732.1" /db_xref="GI:12836591" /translation="MSSSTLSEWPLVAVCOYTPNPKNQENRTCAELVQEARIGACI AFLEADFLARNPAETLLISEPLNGDLGGYSLARPCGIMLSIGGHNEGDMOENOK IKINCVILNSKGSVAAYRKTHICDVEIRSGSNVTKGTEPEPVKIPAGVGAI GVGLIDMRPEPESLIKLAOGAEILTVPSFGSVTGPAHNEVLRLRAIESOCYVIA AACGRHNHTPASGYSHSNVDPMQTVARCEGEGCLARIIDLTFLOMRHLFPQIR RRPDLYSLGNPLPS"
gene	
CDS	
polya_signal	2973..2978

		/note="putative"	
	polyA_site	2993	
		/note="putative"	
BASE COUNT	785 a	739 c	747 g
ORIGIN			722 t
Query Match	62.6%	Score 886.8;	DB 11; Length 2993;
Best Local Similarity	82.6%	Pred. No. 6.9e-140;	
Matches 1123;	Conservative	0;	Mismatches 202; Indels 35; Gaps
OY	42 gatgcagccctgcggaatgttttgctatatatttaagttagagactaccatccc-g	100	
Dd	3 GATCGGACCCCGTGAAATGGTTTGGCTGTGTCATTATAGAACCTACTCCCAAAACCTGG	62	
OY	101 tcggcgcggcctgaggcttcatacacaggccctcctacaagattcctgctcccttgytgc	160	
Dd	63 CCCGCTGGCGCTGGGCTTCATATCACCAGGCCTCTACCAA-----CTCCTGTGTAC	113	
OY	161 tggactccggaatacctaactcaactcagtaacttgytgcacgccagccagaccagcatgyc	220	
Dd	114 CGGATACCGATTAATTTCGAACCCAGTACTTGTACTAGCCCAAGGCCAGAACCATGTC	173	
OY	221 tatctcccttccctcctcgagaactgccccctggtggtctgtgtgccaagttaacctgagcc	280	
Dd	174 ---CTCATCAACTTCTGGGAGCTCCCCCTGGTGGCTGTGTCCAGGTAACTCAACACC	230	
OY	281 agaacagcaaacgaactttaaaacatgacgtgagcttgttcgaagggtctgcagactgg	340	
Dd	231 AATCAACGACAGAGAACTTTAAACATGTGCTGTAGTTGGTTCAAGAGGCTGCCAGACTGGG	290	
OY	341 tgccctgctgagcttccctgctgagcatttgactcatactgcacgggacccctgcagagac	400	
Dd	291 TGCTTGCGCTGGCCTTTCTGGCCTTAGAGCAATTTGACTTTATTCAGCAAAACCTGCGGAGAC	350	
OY	401 gctacacctgtctgacaacaactggtgtgggaaccttttgaagaatacaccagcttgcag	460	
Dd	351 ATTACTCTCTGTCCGAACAACATAAGGGGAGATCTTTGGCCAAATPAGCACCTTGCAG	410	
OY	461 ggaatgtggaactctgctgcttcttggttggttgcatabagctlgaccaagacttggagca	520	
Dd	411 GGAATGTGGAATCTGTGGCTCTCTTGGCGGCTTTCCACAGACGCTGCCCAAGACTGGGAGCA	470	
OY	521 gactcagaanaactlctacaatlgtcaccgtgctgtgacaagcaaaaggagcaagtgtgccac	580	
Dd	471 GAATCAGAANAATCTACAAATTTGTCATGTGCTTTGAACAGCAAGGATGATGTGGCCAG	530	
OY	581 ttacaggaagacacatctgtgtgacgttagagattccagaygcaggggacctatgttgaag	640	
Dd	531 TTACAGGAAGACACATCTGTGCGATGTAGAGATCCAGGCTCGAGGCGCATAGAGAAAG	590	
OY	641 caacctccaatgaccttgagcccaagctcttgaagtcacactltcaacacacagcaagcaaat	700	
Dd	591 CAACATATACCAAGCCTCGAGGCGCATCTTGAGCCACCTGTCAAGAGACCGCGCTGGCAAGT	650	
OY	701 tggctcagctgtctgtcctatgacatbgogtltccctgaactctctcgtgcatactggtcaagc	760	
Dd	651 TGGTCTAGCAATCTGTATAGACATGGGGTTCCTGAACTTCTTGAATTTGGCTCAAGC	710	
OY	761 tggagcagagatacttaactatccttcagacttttggatccattacaagcccaagcccactg	820	
Dd	711 TGGGGGAGAAAATCACTTACTTATCTTCATCCCTTTGGATCTGTTAAGAGTCACGCCACAGC	770	
OY	821 ggaagtggttgcctgggggcccgtgctatgcgaaaaccaaagtctctatgtfagtgagagagaca	880	
Dd	771 GGAGGTGCTGTGCGGCGCCGCGCAATTTGAATCTCAGTGTCTATGTAAATGACAGCGCAG	830	
OY	881 gtgtgagcgcacacatgaaagaagagcaagtatatggccacagcatggtgtgtagaccctgt	940	
Dd	831 GTGTGGAGCCCACTGAAACAAAGAGCAAGTATATGGCCATATACATGNTGTTGACCCGGH	890	
OY	941 gggaacagtgtgtgcccgtctctctgaaggggccagagcctctgcttgccegaataagacct	1000	

Db 391 GGGACAGTGTGGCCGCTGCTCCGAGGAGGCGCTGTGCTTGCATTTGATCT 950
 Qy 1001 caactatcgcagcaagtcgcgcagcaactgcctgtgtccagcaagcgcctgacct 1060
 Db 951 CCACCTTTACACAGATGCGCCACACCTGCTGTTCAGACCGACGACCTACCT 1010
 Qy 1061 ctatgcaatcgtggtcaccacatgctcttaagacttg--actctgtgaattagacctg 1118
 Db 1011 CTATGCACTGTGGTGCATCCACTCTTAATGCTTAACTCTGTGTTAAAGACACA 1070
 Qy 1119 cccctcccccacccacccgcacatagtgctagtgtcatgtgactgtgagggagagac 1178
 Db 1071 GACACAGC-----CAACTGGAGGTGATGCTACTGTGATTTGTAGCAGAGACA 1118
 Qy 1179 cagcagcaagcctccctactctgtgagaacttga-ctctctgtgagaacacagatggct 1237
 Db 1119 CAGAGCAGGTTCCCTTCACTTGCAGAACTTTAACTCTTAAATGSAACACAGGCTTAAT 1178
 Qy 1238 gcttgaggaagaacttcaactcagctcagctcagctcagctcagctcagctcagctcag 1297
 Db 1179 CAAGTCCATGGGAA--AGAAAGAGACTTGTACCTGACCTCAGATTCAGTTTCAGAAAG 1235
 Qy 1298 tggatattatagatcatgttattatattatgaaactgaagttcgtcgtgagcgtgagc 1357
 Db 1236 TGGGATTTATATAGTCACTGTTTATTCATGAAACTGAAGTATGCTGAGGCTGAGC 1295
 Qy 1358 agcactgcagctg---aaaataataatacataagtc 1393
 Db 1296 AGCAGCTGCAATGMAAAAAAATATATATCATATATGTC 1335

RESULT 5
 LOCUS AL520768 890 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL520768 LTI_NFL1004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 5
 prime, mRNA sequence.
 ACCESSION AL520768
 VERSION AL520768.1 GI:12784261
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr.
 FEATURES
 Source
 1..890
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DB002YD05"
 /clone_1ib="LTI_NFL1004_NBC2"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life Technologies. Contact : Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax : (41) 301 610
 8371 Email : fliang@life.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 191 a 251 c 241 g 206 t 1 others
 ORIGIN

Query Match 59.7%; Score 845; DB 9; Length 890;
 Best Local Similarity 96.4%; Pred. No. 1..1e-132;
 Matches 886; Conservative 1; Mismatches 1; Indels 31; Gaps 1;

Qy 20 tctgtctcagaacccgctccctccgagatcgaccccttggaattgttttgctatatctcatg 79
 Db 3 TCTGTCTCCAGACCGCCCTCCGATCGGATCGACCTCGCAAAAGTGTGCTATATCTTCAAT- 61
 Qy 80 taggaactactccatccctcgcgcgcgtgtgagcttcacacagagccctccacag 139
 Db 62 -----GCTGGCGTCAATCACCAGGCTCTCCACAG 91
 Qy 140 attcctgtccctctctgtctcctgtgactccggaatacctcaactctcaacttctgtgtca 199
 Db 92 ATTCTGTCCCTTCTGTGTCGTGAGATCCGGATACCTCAACTCTAGTACTTGTGCTCA 151
 Qy 200 gcccaagcccaagagcaatgtctatctctctctctctctctctctctctctctctctct 259
 Db 152 GCCCAGGCCCAAGGCCATGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 211
 Qy 260 gtgcaggttaacatcgacgcagacagaagaacacttaaaactgtgtcgtgagctgt 319
 Db 212 GTGCCAGGTAACTGACGCCAGACAGACAGAACTTTAAACATGTGCTGAGCTGCT 271
 Qy 320 tcgaagagctgcagagactgtgtgtcctgtgctctctctctctctctctctctctctct 379
 Db 272 TCGAGAGGCTGCGCAGAGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 331
 Qy 380 tgcagggagacctctcagagagcgtacacctgtcttgaaacacatggtgtggaactttgga 439
 Db 332 TGCACGGGACCTGCGAGAGACCTTACCTGCTGTAACACCTGGGGGAACTTTTGA 391
 Qy 440 aagaatacccaagcttgagaggaatgtgactctgtgtctctctctctctctctctctct 499
 Db 392 AGAATACACCACTTGCCAGGGAATGTGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 451
 Qy 500 gcttgcccaagactgtgagcagactcaagaactcaactcaactcaactcaactcaactca 559
 Db 452 GCGTGCCCAAGACTGGGACAGCTCAGAAATCTCAATTTGACAGCTGCTGCTGCAACAG 511
 Qy 560 caaagggcagtaagtgtgccaacttaacagaagaacacatctgtgtgacgtatgaaattccag 619
 Db 512 CAAAGGGCGAGTAGTGCCACTTACAGGAACACATCTGTGTGAGTGAAGATTCCAGG 571
 Qy 620 gcaagggcctatgtgtgaaagaactctacatgctctgtggccagctctgtgactcgt 679
 Db 572 GCAAGGGCTATGTGTGAAGCAACTTACATGCTGTGGCCAGCTTGAGTACACTGT 631
 Qy 680 cagcaaccagcagcagaatgtgtcagctgtctgtctgtctgtctgtctgtctgtctgt 739
 Db 632 CAGCACACGACGACGCAATATTGCTACTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 691
 Qy 740 ctctctgtgcatgtgtcagaactgtgagcagagataacttactctactctctcaagctttggatc 799
 Db 692 CTCTCTGGCATTTGGCTCAAGCTGAGACAGATGACTTACTTACTTCTTCACTTGTGGATC 751
 Qy 800 catcagagcccaagcccaactgtgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 859
 Db 752 CATTTAGGCCCCAGCCACTGTGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 811
 Qy 860 ctatgtagtgagcagcagcagctgtgagcagcagcagcagcagcagcagcagcagcagcag 919
 Db 812 CTATGTAGTGGCAGCAGCAGCAGTGTGAGCCACCATGTGAAGAGAGCAAGTTATGGCCA 871
 Qy 920 cagcatgtgtgttagacc 938
 Db 872 CAGCATGTGTGKAGACCCC 890

RESULT 6
 LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001

DEFINITION	AL529152.LTI.NFL001.NBC4	Homo sapiens	CDNA clone	CS0DD004YM06	5
ACCESSION	prime, mRNA sequence.				
VERSION	AL529152				
KEYWORDS	AL529152.1	GI:12792645			
SOURCE	EST.				
ORGANISM	human.				
	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 844)				
JOURNAL	L.I.W.B., Gruber,C., Jesssee,J. and Polayes,D.				
COMMENT	Full-length cDNA libraries and normalization				
	Unpublished (2001)				
	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				
	BP 191 91006 Evry cedex - France				
	Email: segref@genoscope.cns.fr , Web : www.genoscope.cns.fr .				
FEATURES					
SOURCE	Location/Qualifiers				
	1..844				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0DD004YM06"				
	/clone_id="LTI_NFL001_NBC4"				
	/sex="male"				
	/tissue_type="neuroblastoma cells"				
	/lab_host="DH10B"				
	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA				
	was primed with a NotI-oligo(dt) primer. Five prime end				
	enriched, double-stranded cDNA was digested with Not I and				
	cloned into the Not I and Eco RV sites of the pCMVSPORT 6				
	vector. Library was normalized. Library was constructed				
	by Life technologies. Contact : Feng Liang Life				
	Technologies, a division of Invitrogen 9800 Medical Center				
	Drive Rockville, Maryland 20850, USA Fax : (1) 301 610				
	8373 Email : fliang@lifetechn.com URL :				
	http://fulllength.invitrogen.com "				
BASE COUNT	182 a	238 c	235 g	189 t	
ORIGIN					
Query Match	59.6%; Score 844; DB 9; Length 844;				
Best Local Similarity	100.0%; Pred. No. 1.7e-132;				
Matches 844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY 300	aaacatgtctgagctgtgttcgagagagctccagactgggtgctgtcgtcttcctg	359			
DB 1	AAACATGTGCTGCAGAGCTGTTTCGAGAGGCTGCAGACACTGGGTGCTGCTTCCTG	60			
OY 360	cctgagagcattgacttcattgcacggagccctgcagagaagctacacctgtctgaacca	419			
DB 61	CCTGAGGCAATTGACTTATTCAGAGGAGACCTTCGACAGAGCGCTACACCTGTGTGAACCA	120			
OY 420	ctgggtggaaacttttgaaagataacaccacagcttcgcagggaaatgtgagctcgtgctg	479			
DB 121	CTGGGTGGAAACTTTTGAAAGATAACACCAAGCTTCGACAGGAAATGTGACTCGGCTG	180			
OY 480	tccttgaggtgtttcccaagacgltgccaagagcttggagagagactcaaaaaactacaac	539			
DB 181	TCCTTGAGGTGTTCCTTCATGACGAGCTGTGCCAAGAGCTGGAGACACACTCAGAAAAATCTACAAAT	240			
OY 540	tgtcacgtgtcgtctgaacagcaaaaggagtagtgtgccaacttacagaagaacacatctg	599			
DB 241	TGTCACTGTGCTGTGAACAGCAAAAGGGGCGAGTAGTGCCACACTTACAGGAAGACACATCTG	300			
OY 600	tgtagcgtgaagatccagagggcaggggcctatgtgtgaagaacaaacttaccatgtcgtgg	659			
DB 301	TGTGACGTGAGGATCCAGGGGCGAGGGGCTATGTGTGAAAGCAACTCTACCACTGCTGCGG	360			
OY 660	cccaatctctgagctacacctgtcagcacacaagaaggaagaagatgtgtcagctgtcgtcat	719			
DB 361	CCCACTCTTGAAGTACCTGTGACGACACACAGCAGCAAGATGTTGTCGTACGCTCTCTGCTAT	420			
OY 720	gacatgctgggttcctcgaactctctctgtgcaattgtgtcctaagctgtgagcagaatactaac	779			

Df	421	GACATGGCGTTCCCTGGAACCTCCTCTGTGCATTGGCTCAAGCTGGAGCAGAATACTTACC	
OY	780	tatccctcaagccttttgaataccaattacaaggcccagaaccacactggaaagtgttgtcgcygacc	
Df	481	TATCCTTCAGCTTTTGATCATCATTAAGGCCAGCCACTGGAGAGTGTGCTGGGGCC	
OY	840	cgtgtcatcgaacccccagtgcttatgatgatgcagcagcacagtgtagtcgcgccacatatag	
Df	541	CCTGTATTCGAACAACCACTGCTATGTAACTGTGACAGACAGTAGTGAGACGGCACCATATAG	
OY	900	aagaagaagaattatagccacagacatgtgtgtatagaccttggaagaagtagtgtgccccgc	
Df	601	AAGAAGCAGAAAGTTATGGCCACAGCATGGTGGTAGACCCCCCTGGAGAACAGTGGTGGCCCC	
OY	960	tgctctgaggggcccagagcctctgccccttgccegaataagacctaactatctcgaaagtgtg	
Df	661	TGCTCTGAGGGGGCCAGGGCTCTGCTTGCCTCCGAATATAGACCTCAACTATCTGGAGACATGG	
OY	1020	cgcgcagacactgacctgtgtttccagacacgcagcgcctgacactctatgtgcaatcgtgtaac	
Df	721	CGCCGACACCTGCCCCTGTGTCCAGCACCGCAGAGGCTTGACCTCTATGGCAATCTGGGTGAC	
OY	1080	ccacgtctttaagacttaactctctgtgtgaggttagaccctgcgcccctccacccccaccctgcg	
Df	781	CCACTGCTTTAAGACTTAAGCTTCTGTGTAGATTAGACCTGCCCCCTCCACACCCACCCACTGC	
OY	1140	cact 1143 	
Df	841	CACCT 844	
RESULT	7		
AL562604/C			
LOCUS	AL562604	853 bp	mRNA linear EST 16-FEB-2001
DEFINITION	AL562604 lTL_NFL003_NBC3 Homo sapiens CDNA clone CSDDC015YJ08 3		
ACCESSION	AL562604		
VERSION	AL562604.1 GI:12911188		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1..(bases 1 to 853)		
AUTHORS	Ll.M.B., Gruber,C.; Jesse,J.; and Polayes,D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES			
Source	Location/Qualifiers		
	1..853		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CSDDC015YJ08"		
	/clone_lib="lTL_NFL003_NBC3"		
	/sex="male"		
	/issue_type="neuroblastoma cells"		
	/lab_host="DH10B"		
	/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-Oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	184 a	216 c	238 g
ORIGIN	193 t	22 others	

Db	350	GGTGGCCGCTGTCRCRTAGAGGGGCCAGGGCCTACTGCCTGGCCCCGAATAGACMCMACTAICT	291
OY	1010	-gcagcagtctgcccgcgcacacccctgctgtgttcacagcacgcgaaggccttacctatggca	1068
Dc	290	CGCGCACATTGCGCGCACACCCTGCCTGTGTTCCACACACCGCAGCGCTCACTCATGCGCA	231
OY	1069	atctgggtaccaccaactgcttaagaacttgactttctgtagtatttagactgccctcccaac	1128
Db	230	ATCTGGGTACCCCACTGCTTAAAGCTTGACTTCGTGTAAGTTAAACCCGCCSCCSNAC	171
OY	1129	ccccacctgccaactatagctagctgcatgctcattgacttggagagcgagataccagacagc	1188
Db	170	CCCCACCCCCTACTATAGCTAGTGCCTCANMTGACTTTGGAGGACAGATCCAAGGCACAGC	111
OY	1189	tccctcactctggagaaacttgactctctcttgaatggaacacagatgagctcttggaaaag	1248
Db	110	TCCCCAAACTTGGAGAACCTTGACTCTCTTGTAAGGACAACAGATGGCGCTGGGAANAAG	51
OY	1249	aaacttcaactgagcttccacttgagtgacagctcagtt	1288
Db	50	AAACTTTCMCTGAGCTTCACCTGAGGTGACAGCTGCACTT	11
RESULT	9		
AL563355/c			
LOCUS	AL563355	916 bp	mRNA
DEFINITION	AL563355 LTI_NFL001_NBC4	Homo sapiens	CDNA clone CS0DD004YM06 3
ACCESSION	prime, mRNA sequence.		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 916)		
JOURNAL	L.I.W.B., Gruber,C., Jesse,J. and Polayes,D.		
COMMENT	Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqrel@genoscope.cns.fr, web : www.genoscope.cns.fr. Location/Qualifiers 1..916 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0DD004YM06" /clone_1id="LTI_NFL001_NBC4" /sex="male" /tissue_type="neuroblastoma cells" /lab_host="DH10B" /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand CDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9600 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filiang@lifetech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	227 a	230 c	239 g 202 t 18 others
ORIGIN			
Query Match	53.6%	Score 759.6;	DB 9; Length 916;
Best Local Similarity	96.3%	Pred. No. 2.5e-118;	
Matches	775; Conservative	9; Mismatches 20;	Indels 1; Gaps 1;
589	agacacatctgtgtgacgttagaatccagagcgagggcctatgtgtgaaagcaactcta	648	
1:			

Db	916	ARACACATCTTGTGACGTAGAGATTCCAGGCGACGGGCGCTATGCTGTAAGAACCACTCTA	857
Qy	649	ccatgctctgggcccagttcttgagtcacctgtgcagcacaccagcaggaagaattgctag	708
Db	856	CCATGCTCGGGGCCCGCTTTRAGTCACCTGTGCACACCVGACGACCAAGATGGCTTAG	797
Qy	709	ctgtctgtatgacatgagttcccttgacacctctcttgacttgcttaagtgtgaagag	768
Db	796	CTCTGCTATATACATGCG-GTTCCCTGMACTCTCTGTGGCATTTGGCTCAAGGTGGAGCAG	738
Qy	769	agatactaccatctcttcacagcttttgatccattacagagcccaccccaacttgagagtg	828
Db	737	AGATACCTTACCTATTCCTTACAGTTTGGATATCATTAACAGCCGACCCACTGAGAGTGCT	678
Qy	829	tgtctggggcccggtgctatcgaaccacagtgctatgtatgtgtagcagcacagatgtgac	888
Db	677	TGCTCGGGGCCCGTGTCTATGCAAAACCCAGTGTATGTAGTGACACAGACAGTGTGAC	618
Qy	889	gccacacatgagagaagagcaagttatctgtgccaaagacaatgtgtgttaaacctctgggaaacag	948
Db	617	GCCACCATGAGAGAAAGACCAAGTTATGACCTTTGATGATGTGTAAACCCCTGGGAAACAG	558
Qy	949	tgggtggccgcgtgcgtctagaggccagggccctgtgcctctggccgaaataacctcaactac	1008
Db	557	TGTGTGGCCCGCTGCTCTGTAGGGGCCAGGCGCTCTGCTTGCCCGAATTAACCTCAACTTTC	498
Qy	1009	tgcgacagttgcgcgcgacacactgcctgtgttctcagacacgcgcagcctctgacctctatg	1068
Db	497	TGCGACAGTGTGGCGCGACACCTGCTGTATGTTCCACACCGCAGGCGCTGACCTTATMGGA	438
Qy	1069	atctgtgggtcaaccactgtctcttaagaacctgtactctgtgtatgttagacctgtccctccac	1128
Db	437	ATCTGGGTACACCCACTGTCTTAAAGACTTACTTCTGTAGATTAAACCTGCCCCCTCCSCC	378
Qy	1129	cccccaactgcgcacatbagcttagtgcctcatgttgactctggaagcagagatccagcgacacag	1188
Db	377	CCCTBCCCTGCACCTTGTAGCTTAGGGCTCAATGTGACTTTGGAGGACAGATCCAGGCACCGC	318
Qy	1189	tcacctcaacttgagaaaccttgacctctctctgtatggaaacaacagatbggctgtcttggaaag	1248
Db	317	TCCGNNNNBTGGAGAACCTTGACTTCTGTATGGAAACAGATAGGCGCTTGGGAAAG	258
Qy	1249	aaacttcaacctgagcttcaacctgtagctcagactgtcagttcagaagaagtgtgaatttat	1308
Db	257	ANACTTTCACCTGACCTTCACTCAGCTGAGGTCAGACTCAGTTTCAGAAGGTGGAATTTTAT	198
Qy	1309	atagtcattgtttattattatggaaactgaagtctgcctggaagcctggaagacagcggcat	1368
Db	197	ATAGTCATTGTTTATTTTATGTGAAGCTGAAGTTGTGTGAGGGCTGTAGCAGNACTGGCAT	138
Qy	1369	tgaataataataatacatcaataagtc	1393
Db	137	TGAAAAATATATATATCATTAAGTC	113
RESULT	10		
LOCUS	BI769604	846 bp	linear
DEFINITION	60305498571 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',		EST 25-SEP-2001
ACCESSION	BI769604		
VERSION	BI769604.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-remail.nih.gov		

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11512 row: j column: 22
 High quality sequence stop: 838.

FEATURES

source

Location/Qualifiers
 1. 846

/organism="Homo sapiens"
 /db_xref="taxon:9606"

/clone_1b="IMAGE:5204469"
 /clone_1lb="NIH_MGC_1122"

/lab_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: pCMV-Sport6;
 Site_1: NotI; Site_2: EcoRV (destroyed); RNA source

anonymous pool of 24 week female lung, 16 week female
 spleen, and 20-22 week male spleens. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.4 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 026. Note:
 this is a NIH_MGC Library."

BASE COUNT 179 a 240 c 226 g 201 t
 ORIGIN

Query Match

53.5%; Score 758; DB 10; Length 846;

Best Local Similarity 96.0%; Pred. No. 4.7e-118;

Matches 843; Conservative 0; Mismatches 0; Indels 35; Gaps 5;

QY 20 tcgtggtcgaagccggccctccgagatcgagccctcggaatggttggtgatatcttcacg 79
 Db 2 tcgtggtcgaagccggccctccgagatcgagccctcggaatggttggtgatatcttcacg 60
 QY 80 taagaaactactccctacccgctgcgcgcgcgtgagcttcacacagagccctccacag 139
 Db 61 -----GCTGGGCTTCATCACCAGGCTCTCCACAG 90
 QY 140 attcctgtccctctgtgtctcgtgacctcgagatacctcaactctcagttctgtctca 139
 Db 91 attcctgtccctctgtgtctcgtgacctcgagatacctcaactctcagttctgtctca 150
 QY 200 gcccaagcccaagagcgaatgctatcctcctctcctcctcgcgaactccctcgtgagctg 259
 Db 151 gcccaagcccaagagcgaatgctatcctcctcctcctcctcgcgaactccctcgtgagctg 210
 QY 260 gtgcagagtaacatcgacgcagacagacaagaacttaaacatgtgtgtaagctgtg 319
 Db 211 gtgcagagtaacatcgacgcagacagacaagaacttaaacatgtgtgtaagctgtg 270
 QY 320 tcgagagctgtccagacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 379
 Db 271 tcgagagctgtccagacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 330
 QY 380 tcgagagctgtccagacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 439
 Db 331 tcgagagctgtccagacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 390
 QY 440 agaatacacccagctgtccagaggaatgtgactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 499
 Db 391 agaatacacccagctgtccagaggaatgtgactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 450
 QY 500 gcgtgtgccaagactgt 559
 Db 451 gcgtgtgccaagactgt 510
 QY 560 caaagggagcagtagtggt 619
 Db 511 caaagggagcagtagtggt 570

QY 620 gcaagggcctatgtgtgaaagcaactctacatcctcgtggccagctgtgactcctgt 679
 Db 571 gcaagggcctatgtgtgaaagcaactctacatcctcgtggccagctgtgactcctgt 630
 QY 680 cagacacacagcagcagaatgtgtcctgtgtc-tgtatgacatgtgtgtgtgtgtgtgt 738
 Db 631 cagacacacagcagcagaatgtgtcctgtgtc-tgtatgacatgtgtgtgtgtgtgtgt 690
 QY 739 tcctctgtgcatgtgtcctcagctgtgagcagagatcttaccctcctcagctgtgtgt 798
 Db 691 tcctctgtgcatgtgtcctcagctgtgagcagagatcttaccctcctcagctgtgtgt 750
 QY 799 ccattacagccagccacccactgtggaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 858
 Db 751 ccattacagccagccacccactgtggaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 808
 QY 859 gctatgtatgtgtgtgt-cagcacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 895
 Db 809 gctatgtatgtgtgtgt-cagcacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 846

RESULT 11

LOCUS

AL559163 861 bp mRNA linear EST 16-FEB-2001

DEFINITION

AL559163 LTI_NFL008.TC2 Homo sapiens cDNA clone CS0DJ014YA17 5

ACCESSION

AL559163

VERSION

AL559163.1 GI:12904391

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 861)

AUTHORS

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr.

FEATURES

source

Location/Qualifiers

1. 861

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_1b="CS0DJ014YA17"

/clone_1lb="LTI_NFL008.TC2"

/sex="male"

/tissue_type="T cells from T cell leukemia"

/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact: Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax: (1) 301 610 8371

Email: fliang@lifetech.com URL: <http://fulllength.invitrogen.com>

BASE COUNT 176 a 229 c 225 g 201 t 30 others

ORIGIN

Query Match

53.5%; Score 758; DB 9; Length 861;

Best Local Similarity 95.0%; Pred. No. 4.7e-118;

Matches 790; Conservative 29; Mismatches 9; Indels 4; Gaps 4;

QY 110 gctgggtcattacacagcctcctacagatcctgtcctcctgtgtgtgtgtgtgtgtgtgt 169
 Db 30 gctgggtcattacacagcctcctacagatcctgtcctcctgtgtgtgtgtgtgtgtgtgt 89
 QY 170 gataccctaacctcagttgt 229

Db	640	TGTCMGATAGACATAGCGGTTCCTCCGAACTCCTCTGGCATTTGGCTCAAGCTGGAGACAGA	699
Oy	770	gatacttacc-tatccttcagcttttgatccattaca---ggcccagcccaetgyagg	825
Db	700	GATACTTACCTTATCTTACCTTCAGCTTTTGATTCCTAATACAGGCCAGCCCACTGGGAAG	759
Oy	826	tgttgtcctggggccccgtgctatcgaaacccagtgtatg-tgttggcagcacacaagyt	884
Db	760	TGTTCTGTGGGGNCCGTCTATTGCAGAACCCAATGCTATGTGTGTGGCCGACACAGTGT	819
Oy	885	gg-agccaccactgagaagagacgaatgatgcacacagcatgctgtagtacctgag	943
Db	820	GSAAAGCCACCATGTAGAGAGAGAGC-AGTTATGGGCACACAGATGCTTAACCTTGGGG	878
Oy	944	aacagttg 951	
Db	879	ACAGGGGG 886	
 RESULT_13			
BG436916		793 bp	mRNA linear EST 14-MAR-2001
LOCUS	BG436916	602488424F1 NIH_MGC_18 Homo sapiens CDNA clone IMAGE:4620892 5'	
DEFINITION		mRNA sequence.	
ACCESSION	BG436916		
VERSION	BG436916.1	GI:13343422	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 793) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
AUTHORS	Contact: Robert Strausberg, Ph.D.		
TITLE	Email: cga@bs-remai.nih.gov		
JOURNAL	Tissue Procurement: DCTD/DTP/Gazdar		
COMMENT	CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLB) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB at: http://image.lnl.gov Plate: LINCMI382 row: O column: 05 High quality sequence stop: 767. Location/Qualifiers 1. 793 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4620892" /clone_lib="NIH_MGC_18" /tissue_type="large cell carcinoma" /lab_host="DH10B (phage-resistant)" /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA Synthesis Kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
FEATURES			
SOURCE			
BASE COUNT	181 a 210 c 230 g 172 t		
ORIGIN			
Query Match	51.2%; Score 724.4; DB 10; Length 793;		
Best Local Similarity	98.0%; Pred. No. 2.le-112;		
Matches 776; Conservative 0; Mismatches 11; Indels 5; Gaps 4;			
Oy	242	actgcccttggtgctgtgtcgcaggtatcacatgcagccagacagaacagaaactttaa	301
Db	2	ACTGCCCTGTGGTGTGTGTGCAGTAGATCGAGCGCCAGACAAGCAAGAACTTTAA	61

QY	302	aacatgtctgagcgtgatttcggaagcctgcccagacttgggtgcctctgaccttccctgc	361
Db	62	AACATGTGCTGAGCTGGTTCCGAGAGGCTGCCAGACTGGGTGCTGCTTCTCTGGC	121
QY	362	tgaagcatttgatcttcattatgcaacgagacccttgacaagacgtctacacctgtctgaacact	421
Db	122	TGAGGCATTGTGACTTCATTGCGACGGGACCTCTCAGAGAGCGTACACCTGTGAACT	181
QY	422	gggttgggaacttttgggaagaaataccacagcttgcacaggaaattgtgacttgcctgc	481
Db	182	GGGTGGGAAACTTTTGGAAAGATTAACACCAGCTTCCAGGGAAATGTGACTGTGGCTGTC	241
QY	482	cttgggttgatttccataagagcctgctgcacaagacgtggagacagactcaagaaatctacaatg	541
Db	242	CTTGGGTGGTTTCCATGACGCTGGCCAGACACTGGGAGCAGACTCAGAAAATCTACAAATGG	301
QY	542	tcaagtgctcgtctgaacaagcaaaagggtgaagtgagccacttaagaagaacacatctgtg	601
Db	302	TCACGTGCTGTGCGAAGACACCAAGGGGGGAGTAAGTGGCCACTTACAGGAAGACACATCTGTG	361
QY	602	tgaagtaagatcttcagggtcagggtccatgtgtgaagcaacttaccatgctcctgggccc	661
Db	362	TGACGTAAAGATTCACGGGCGAGGGGCTCATGTGTAAAGCAACTTACCATCTCCGTGGGCC	421
QY	662	caagcttgaagtaacctgtcaagcacacacacagcaagaaatgtgtcagctgtctgtatga	721
Db	422	CAGTCTTGAAGTACACTGTGACGACACACAGGCAAGATGGTCTTAGCTGTGCTATGA	481
QY	722	catgcggtcccttgaaactctctctgtgcatltgtctaaagcttgaagacagataacttaacta	781
Db	482	CATGGGTTTCCCTGAACTCTCTGTGGCATTTGGCTCAACCTGGAGCAGAGATCTTACCTA	541
QY	782	tcctt-caagcttlttgatccaatlaaagcccaagcccacttgggaagtgtgtctgaggg-cc	839
Db	542	TCTTACAGGCTTTTGATTCATTACAGGCCCGCCACCTGGGAGGTGTGTCTGCGGGCCC	601
QY	840	ctgtgtatcgaaacacagctgctatgtlaqtgagcagcacagctgtgtgaagccacacatgag	899
Db	602	CGTGTATTCGAAGAACCCATGCTGCTATGTAGTGGCAGCAGCACAATGTGTGAGCGCACCATGAG	661
QY	900	aagagagcaagttatlgcccacagacatggtgtgtagacccc-ttggggaacagttgtgtgccc	957
Db	662	AAGAAGCAAGATTATGCGCCACAGCATGTGTGTAGAACCCCTTGGGGAAACAGATGTGGGCC	721
QY	958	gctgtcttgaggggccaagccctctgtccttg-ccgaatagactctaactatcttggaag	1016
Db	722	GCTGTCTTGAAGGGGCGCAGGCTCTCTGCTTGGCCCGAATAGACTCTCAAGATCTTGGCAGG	781
QY	1017	ttgcgcgcacac 1028	
Db	782	TTGCGCCGCAAC 793	
RESULT 14			
LOCUS	Bg703176	790 bp	mrna linear EST 07-MAY-2001
DEFINITION	602686082p1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4818405 5',		
ACCESSION	Bg703176		
VERSION	Bg703176.1	GI:13975252	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 790)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapds-remail.nih.gov		
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		

Best Local Similarity 93.4%; Pred. No. 4.8e-109;
Matches 735; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
QY 615 ccaaggcaggggctatgtgtgaaagcaactctccatgacctggtgcccagcttgagtc 674
    || || || || || || || || || || || || || || || || || || || || ||
Db 787 CCCAGGGCAGGGCCTAATGTTGAAAGCAATTCGCCAGGCTTGGCCCGCTGTGTCA 728

QY 675 ccttcagcacacagcaggaagaattgtctagctgtctgtatgacatgcggtccct 734
    || || || || || || || || || || || || || || || || || || || || ||
Db 727 CCGGTCACTACACCAACGCTAAGTGTGTGTACGTGTTCGAATGACATCGGTTCCCTC 668

QY 735 gaactctctcgtcattggtctaaagcttgagcagaagatacttaacctcctcagctt 794
    || || || || || || || || || || || || || || || || || || || || ||
Db 667 GAACCTCTCTCGGCAATGGCTCAAGCTGGACAGAACTACCTCATCTTCAGCTTTG 608

QY 795 ggaaccattlaagggccagccacttgaggtgtgtctgcgggcccgtcatgaaacc 854
    || || || || || || || || || || || || || || || || || || || || ||
Db 607 GCATCCATCACAGGCCCGCCACTGGGAGGTGTCTGGGGCCCGCTGTATCGAAAC 548

QY 855 cagtgctatgtatgtgacagcagcaagtgtagcagccacatgaagaagagcaagttat 914
    || || || || || || || || || || || || || || || || || || || || ||
Db 547 CAGTGCATATGATGTGGCAGCGCCAGTGTACGCCCATGAGAAAGAGCCAGTTAT 488

QY 915 ggcacacagcatgtgtgtagacccttgggaaacagtggtgcgcgtgctctgagggcca 974
    || || || || || || || || || || || || || || || || || || || || ||
Db 487 GGCACACAGCATGTGTAGACCCCTGGGGAAACAGTGTGGCCCGCTGTGAGGGGCCA 428

QY 975 ggcctctgcctccgcaatagactaactatctgacaggttgccgcaactgct 1034
    || || || || || || || || || || || || || || || || || || || || ||
Db 427 GGCCTGTGCTTGGCCCGAATGACCTCAACTGTGCGACAGTTGGCGGACCTGCT 368

QY 1035 gtgtccagcacagcagcctgcactctatgtgcaatctgtgtccaccactgtctaaagac 1094
    || || || || || || || || || || || || || || || || || || || || ||
Db 367 GTGTTCCAGACGCCAGGCCCTGACCTCTATGCAATCTGGGTCAACCCACTGTTTAAAGAC 308

QY 1095 ttgactctgtgagtttagacctgccccctccacccccacccctgccaactatgagctagtg 1154
    || || || || || || || || || || || || || || || || || || || || ||
Db 307 TTGACTTCTGTGAGTTTAGACNTGTGCCCTCCACCCCGCTGCACTATGAGCTAGTG 248

QY 1155 ctcaatgtgacttgagagcagatccagcagacagctccctacttgagaaacctgaactc 1214
    || || || || || || || || || || || || || || || || || || || || ||
Db 247 CTCATGTGACTTGGAGGAGGATCCAGGCACAGCTCCCCCTCATTGGAGAACTTGAAGCTC 188

QY 1215 tcttgatggaacacagatggctgtcttggaagaanaacttcacctgagcttcaacctgag 1274
    || || || || || || || || || || || || || || || || || || || || ||
Db 187 TTTTGATGGAACACAGATGGGCTGCTTGGAAGAACTTTCACCTGAGCTTCACCTGAG 128

QY 1275 gtccagactcagtttcagaaagtggaatttatatagtcattgttatttcatggaac 1334
    || || || || || || || || || || || || || || || || || || || || ||
Db 127 GTCAGACTGCAAGTTTCAGAAAGGTGGAATTTATATAGTCATGTGTTTTCATGGAAC 68

QY 1335 tgaagttctgtgagggctgagcagccactgcatgaaanaataataataaagfca 1394
    || || || || || || || || || || || || || || || || || || || || ||
Db 67 TGAAGTTCTGTGAGGGCTGAGCAGCCCTGGCATTGAAATAATATATATCAATAAGTCT 8

QY 1395 aaaaaa 1401
    || || || || || || || || || || || || || || || || || || || || ||
Db 7 GAAAAAA 1
```

Search completed: April 28, 2002, 04:17:32
Job time: 10007 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2002, 01:43:40 ; Search time 52.84 Seconds

(without alignments)
6562.457 Million cell updates/sec

Title: US-09-357-675C-1

Perfect score: 1416
1 gccacacgcgtgcgcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Sequence: 1 gccacacgcgtgcgcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCUS.COMB.seq:*
6: /cgn2_6/ptodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	3.3	7218	1	US-08-232-463-14 Sequence 14, Appl
2	41.2	2.9	1332	4	US-09-333-423-1 Sequence 1, Appl
3	41.2	2.9	1602	1	US-08-530-950-3 Sequence 3, Appl
4	41.2	2.9	1602	3	US-08-888-429A-3 Sequence 3, Appl
5	41.2	2.9	1602	4	US-09-149-879-3 Sequence 3, Appl
6	40.4	2.9	1505	4	US-09-385-982-262 Sequence 262, App
7	40.4	2.9	1505	2	US-08-909-965C-13 Sequence 13, Appl
8	39.8	2.8	10660	2	US-08-267-803B-8 Sequence 8, Appl
9	39.8	2.8	10660	4	US-09-041-886-16 Sequence 16, Appl
10	39.2	2.8	1721	5	PCR-US96-00994-3 Sequence 4, Appl
11	39.2	2.8	3982	3	US-08-947-823-4 Sequence 4, Appl
12	38.8	2.7	208	1	US-08-686-878A-37 Sequence 37, Appl
13	38.8	2.7	208	4	US-09-175-928-37 Sequence 37, Appl
14	38.8	2.7	911	2	US-08-924-759-9 Sequence 9, Appl
15	38.8	2.7	911	3	US-09-248-335-9 Sequence 9, Appl
16	38.8	2.7	1393	1	US-07-602-824A-1 Sequence 1, Appl
17	38.8	2.7	1393	1	US-07-983-451-1 Sequence 1, Appl
18	38.8	2.7	1393	1	US-08-261-577-6 Sequence 1, Appl
19	38.6	2.7	2550	6	5258287-23 Patent No. 5258287
20	38.2	2.7	2852	3	US-09-027-137-2 Patent No. 5169941-7
21	37.8	2.7	1215	6	US-09-232-200-68 Patent No. 5169941
22	37.8	2.7	1998	4	US-09-232-197-68 Sequence 68, Appl
23	37.8	2.7	1998	4	US-09-232-201-68 Sequence 68, Appl
24	37.8	2.7	1998	4	US-09-232-191-6 Sequence 6, Appl
25	37.8	2.7	2087	4	US-09-232-200-6 Sequence 6, Appl
26	37.8	2.7	2087	4	US-09-232-197-6 Sequence 6, Appl
27	37.8	2.7	2087	4	US-09-232-197-6 Sequence 6, Appl

28	37.8	2.7	2087	4	US-09-232-201-6 Sequence 6, Appl
29	37.6	2.7	1910	4	US-08-974-691-7 Sequence 7, Appl
30	37.6	2.7	2239	4	US-09-196-390-1 Sequence 1, Appl
31	37.4	2.6	308	2	US-08-721-488-4 Sequence 4, Appl
32	37.4	2.6	536	1	US-08-341-568-1 Sequence 1, Appl
33	37.4	2.6	536	2	US-08-911-020-1 Sequence 1, Appl
34	37.4	2.6	2589	4	US-08-569-749-1 Sequence 1, Appl
35	37.4	2.6	2589	5	PCR-US96-12860-1 Sequence 1, Appl
36	37.2	2.6	90	1	US-08-677-944-1 Sequence 1, Appl
37	37.2	2.6	90	4	US-09-254-048A-1 Sequence 1, Appl
38	37.2	2.6	991	3	US-08-924-747-25 Sequence 25, Appl
39	37.2	2.6	991	4	US-09-247-373B-25 Sequence 25, Appl
40	37.2	2.6	991	4	US-09-296-715-25 Sequence 25, Appl
41	37	2.6	570	1	US-07-885-970A-10 Sequence 10, Appl
42	37	2.6	570	1	US-08-298-687A-10 Sequence 10, Appl
43	37	2.6	570	1	US-08-298-829-10 Sequence 10, Appl
44	37	2.6	599	4	US-09-328-111-147 Sequence 147, App
45	37	2.6	609	1	US-08-530-797-9 Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ. ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match 3.3%; Score 47; DB 1; Length 7218;
Best Local Similarity 1.9%; Pred. No. 0.00064;
Matches 5; Conservative 164; Mismatches 95; Indels 0; Gaps 0;

QY 2 cccactgcgtgcgctnctgtctccagaccgcccctccgcatgcgacctcgcaatggt 61
DB 1177 yyy 1236
QY 62 ttctgcatatcttcatgaagaccctactccctaccgctcgccgagctggctcat 121
DB 1237 yyy 1296
QY 122 caccagcctcctcacaagctctcctctgtctctgactccgatacctaact 181
DB 1297 yyy 1356
QY 182 cccactacttctgtctcagcccgccgagccatgctctctctctctctcgca 241
DB 1357 yyy 1416
QY 242 actgccccctgctgctgtgcca 265
DB 1417 yyyyyyyyyyyyyyytaccga 1440

RESULT 2
US-09-333-423-1
Sequence 1, Application US/09333423
Patent No. 6265636
GENERAL INFORMATION:
APPLICANT: Randall, Douglas
APPLICANT: Thelen, Jay
APPLICANT: Miernyk, Jan
APPLICANT: Muszynski, Michael
APPLICANT: Sewalt, Vincent
TITLE OF INVENTION: Pyruvate Dehydrogenase Kinase
TITLE OF INVENTION: Polynucleotides, Polypeptides and Uses Thereof
FILE REFERENCE: 0818
CURRENT APPLICATION NUMBER: US/09/333,423
CURRENT FILING DATE: 1999-06-15
EARLIER APPLICATION NUMBER: 60/089,998
EARLIER FILING DATE: 1998-06-19
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1332
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (55)...(1095)
US-09-333-423-1

Query Match 2.9%; Score 41.2; DB 4; Length 1332;
Best Local Similarity 64.9%; Pred. No. 0.014;
Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1323 ttctatggaactgaagctctgtcgaaggctgaagcagcactgcatgtgaaataata 1382
DB 1239 tttgatctgcatctcctaactcgcttattgactgattgtcaaaaaaaaaaaaaa 1298
QY 1383 atcataagtcataaaaaaaaaaaaaaaaaaaaaa 1416
DB 1259 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1332

RESULT 3
US-08-530-950-3
Sequence 3, Application US/08530950
Patent No. 5736381
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.

APPLICANT: Raingeaud, Joel
APPLICANT: Gupta, Shashi
APPLICANT: Derjard, Benoit
TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
TITLE OF INVENTION: KINASES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,950
FILING DATE: 19-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1602 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-530-950-3

Query Match 2.9%; Score 41.2; DB 1; Length 1602;
Best Local Similarity 57.9%; Pred. No. 0.016;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1291 agaaagtggaattatatagctgttattctcaggaacagtcgaagtcgtcgag 1350
DB 1444 AGATGATGATTAATGATTAAGGCTTAGACTTCAAAAGTGATTTAATTAATGATG 1503
QY 1351 gctgagcagcactggtgcatgtaaaataataataataaagtcacaaaaa 1410
DB 1504 TGTCAATGAGTCCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1563
QY 1411 aaaaaa 1416
DB 1564 AAAAAA 1569

RESULT 4
US-08-888-429A-3
Sequence 3, Application US/08888429A
Patent No. 6136596
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
APPLICANT: Whitmarsh, Alan
APPLICANT: Tournier, Cathy
TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-
TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE KINASES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA

```

: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/888,429A
: FILING DATE: 07-JUL-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/530,950
: FILING DATE: 19-SEP-1995
: APPLICATION NUMBER: 08/446,083
: FILING DATE: 19-MAY-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Fasse, Peter J.
: REGISTRATION NUMBER: 32,983
: REFERENCE/DOCKET NUMBER: 07917/053001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 299354
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1602 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 244...1245
: US-08-888-429A-3
:
: Query Match
: Best Local Similarity 57.9%; Score 41.2; DB 3; Length 1602;
: Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

: QY 1291 agaaagtggaatttattatagctgttatttcattggaactgaagtcgtcgaag 1350
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 1444 AGATGATGATTAATGATTAAGCTTAGACCTCAAAAGTGATTAATTTAATGATG 1503
:
: QY 1351 gctgagcagcactgcactgtaaaataataatacgaagtcacaaaaa 1410
:   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1504 TGTCATATGATCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1563
:
: QY 1411 aaaaaa 1416
:   |||||
: Db 1564 AAAAAA 1569
:
: RESULT 5
: US-09-149-879-3
: Sequence 3, Application US/09149879
: Patent No. 6174676
: GENERAL INFORMATION:
: APPLICANT: Davis, Roger J.
: APPLICANT: Raingaud, Joel
: APPLICANT: Gupta, Shashi
: APPLICANT: Derjard, Benoit
: TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
: TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/149,879
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/530,950
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Fasse, J. Peter
: REGISTRATION NUMBER: 32,983
: REFERENCE/DOCKET NUMBER: 07917/010001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1602 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-149-879-3
:
: Query Match
: Best Local Similarity 57.9%; Score 41.2; DB 4; Length 1602;
: Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

: QY 1291 agaaagtggaatttattatagctgttatttcattggaactgaagtcgtcgaag 1350
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 1444 AGATGATGATTAATGATTAAGCTTAGACCTCAAAAGTGATTAATTTAATGATG 1503
:
: QY 1351 gctgagcagcactgcactgtaaaataataatacgaagtcacaaaaa 1410
:   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1504 TGTCATATGATCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1563
:
: QY 1411 aaaaaa 1416
:   |||||
: Db 1564 AAAAAA 1569
:
: RESULT 6
: US-09-385-982-262/c
: Sequence 262, Application US/09385982
: Patent No. 6262334
: GENERAL INFORMATION:
: APPLICANT: ENDEGE, WILSON O., ET AL.
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: FILE REFERENCE: CODNA-260XX
: CURRENT APPLICATION NUMBER: US/09/385,982
: CURRENT FILING DATE: 1999-08-30
: EARLIER APPLICATION NUMBER: 09/328,111
: EARLIER FILING DATE: 1999-06-08
: EARLIER APPLICATION NUMBER: 60/117,393
: EARLIER FILING DATE: 1999-01-27
: EARLIER APPLICATION NUMBER: 60/098,639
: EARLIER FILING DATE: 1998-08-31
: NUMBER OF SEQ ID NOS: 144
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 262
: LENGTH: 593
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(593)
: OTHER INFORMATION: n = A,T,C or G
```


QY 1286 gtctcagaagtggaatttatatagtcattgtttatctcatggaactgaagttctgc 1345
|||||
Db 10523 GTTTTAAATTAATTTCTGTATACCACTTATTTGTTTCAGTATTAATCTGTACTAA 10582
QY 1346 tgaaggctgagcagcactgcatgtgaataataataatcataatcaaaataaaaaa 1405
|||||
Db 10583 TAAATTAACAGTCCCAATTGCCAAAAAATAAAAAAAAAAAAAAAAAAAAAA 10642
QY 1406 aaaaaaaaaa 1416
|||||
Db 10643 AAAAAAAAAA 10653

RESULT 9
US-09-041-886-16
Sequence 16, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabizadeh, Sharrooz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 10660 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 936..3384
US-09-041-886-16

Query Match 2.8%; Score 39.8; DB 4; Length 10660;
Best Local Similarity 56.5%; Pred. No. 0.12; Mismatches 57; Indels 0; Gaps 0;
Matches 74; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1286 gtctcagaagtggaatttatatagtcattgtttatctcatggaactgaagttctgc 1345
|||||
Db 10523 GTTTTAAATTAATTTCTGTATACCACTTATTTGTTTCAGTATTAATCTGTACTAA 10582
QY 1346 tgaaggctgagcagcactgcatgtgaataataataatcataatcaaaataaaaaa 1405
|||||
Db 10583 TAAATTAACAGTCCCAATTGCCAAAAAATAAAAAAAAAAAAAAAAAAAAAA 10642
QY 1406 aaaaaaaaaa 1416
|||||
Db 10643 AAAAAAAAAA 10653

Db 10643 AAAAAAAAAA 10653
|||||
RESULT 10
PCT-US96-00994-3
Sequence 4, Application us/
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: SELF-RENEWING PLURIPOTENT HEMATOPOIETIC
TITLE OF INVENTION: STEM CELL COMPOSITIONS, METHODS OF USE,
TITLE OF INVENTION: AND CULTURE SYSTEMS THEREFOR
NUMBER OF SEQUENCES: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US SN 08/462,108
FILING DATE: 05-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US SN 08/378,144
FILING DATE: 24-JAN-1995
Sequence 3, Application us/
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: SELF-RENEWING PLURIPOTENT HEMATOPOIETIC
TITLE OF INVENTION: STEM CELL COMPOSITIONS, METHODS OF USE,
TITLE OF INVENTION: AND CULTURE SYSTEMS THEREFOR
NUMBER OF SEQUENCES: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US UNKNOWN
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US SN 08/462,108
FILING DATE: 05-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US SN 08/378,144
FILING DATE: 24-JAN-1995
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 63..1436
PCT-US96-00994-3

Query Match 2.8%; Score 39.2; DB 5; Length 1721;
Best Local Similarity 66.7%; Pred. No. 0.066;
Matches 56; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1333 actgaagttctgctgagggctgagcagcactgcatgtgaataataataatcataaagt 1392
|||||
Db 1637 ATTAAGATCTTTTGTAGGAAATTAAGAAAAAGAAAGAAAAAATAAAAAAAAAA 1696
QY 1393 caaataaaaaaaaaaaaaaaaaa 1416
|||||
Db 1697 AAAAAAAAAAAAAAAAAA 1720

RESULT 11
US-08-947-823-4

Sequence 4, Application US/08947823
Patent No. 6114605
GENERAL INFORMATION:
APPLICANT: Williamson, Valerie M.
APPLICANT: Kaloshian, Isgonhi
APPLICANT: Yaghoobi, Jafar
APPLICANT: Bodeau, John
APPLICANT: Milligan, Stephen
TITLE OF INVENTION: Procedures and Materials for Conferring
TITLE OF INVENTION: Pest Resistance in Plants
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/947,823
FILING DATE: 09-OCT-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18802
FILING DATE: 09-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-070210US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 87..3860
OTHER INFORMATION: /note= "Copy 2 CDNA for M1 nematode"
US-08-947-823-4

Query Match 2.8%; Score 39.2; DB 3; Length 3982;
Best Local Similarity 56.1%; Pred. No. 0.1;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY 1280 actgcagttcgaagaagtgaatttatatagtcatttattcatggaactgaag 1339
DB 3851 ATTTAAGTGAATTTGGTGAACCTTGCTGTGATATGATGAATTAATAATCCG 3910
OY 1340 tctcgtgagggcgtgacgacgtgcattggaataataataataagaacaaanaa 1399
DB 3911 GTGATGAGATTCTCTTAATTTCTTTTAACAAAATAATATTTTATTAAGTACAAA 3970
OY 1400 aaaaaaaaaa 1411
DB 3971 AAAAAAAAAA 3982

RESULT 12

US-08-686-878A-37
Sequence 37, Application US/08686878A
Patent No. 5708157
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallee, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/686,878A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 208 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-686-878A-37

Query Match 2.7%; Score 38.8; DB 1; Length 208;
Best Local Similarity 54.0%; Pred. No. 0.028;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 1304 ttatataagtcattgtttatcatggaactgaagtcctcgtgaggtcgcagcact 1363
DB 73 TTATATTATTCCTCCNGTAATTTGTCGNGNGATRAATTGCAANTTAACNGCATTAANGNT 132
OY 1364 ggcattgaaataataataataataataataataataataataataataata 1416
DB 133 NATGNTAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 185

RESULT 13
US-09-175-928-37
Sequence 37, Application US/09175928A
Patent No. 6312921
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice

APPLICANT: M1, Sha

APPLICANT: Genetics Institute, Inc.
 FILE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
 CURRENT APPLICATION NUMBER: US/09/175,928A
 CURRENT FILING DATE: 1998-10-20
 NUMBER OF SEQ ID NOS: 62
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 37
 LENGTH: 208
 TYPE: DNA
 ORGANISM: Mus musculus
 FEATURE: NAME/KEY: unsure
 LOCATION: (8)
 FEATURE: NAME/KEY: unsure
 LOCATION: (29)
 FEATURE: NAME/KEY: unsure
 LOCATION: (31)
 FEATURE: NAME/KEY: unsure
 LOCATION: (42)
 FEATURE: NAME/KEY: unsure
 LOCATION: (55)
 FEATURE: NAME/KEY: unsure
 LOCATION: (65)
 FEATURE: NAME/KEY: unsure
 LOCATION: (75)
 FEATURE: NAME/KEY: unsure
 LOCATION: (86)
 FEATURE: NAME/KEY: unsure
 LOCATION: (91)
 FEATURE: NAME/KEY: unsure
 LOCATION: (98)
 FEATURE: NAME/KEY: unsure
 LOCATION: (100)..(101)
 FEATURE: NAME/KEY: unsure
 LOCATION: (113)
 FEATURE: NAME/KEY: unsure
 LOCATION: (119)
 FEATURE: NAME/KEY: unsure
 LOCATION: (121)
 FEATURE: NAME/KEY: unsure
 LOCATION: (128)
 FEATURE: NAME/KEY: unsure
 LOCATION: (130)..(131)
 FEATURE: NAME/KEY: unsure
 LOCATION: (133)
 FEATURE: NAME/KEY: unsure
 LOCATION: (137)
 FEATURE: NAME/KEY: unsure
 LOCATION: (161)
 US-09-175-928-37

Query Match 2.7%; Score 38.8; DB 4; Length 208;
 Best Local Similarity 54.0%; Pred. No. 0.028;

Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 QY 1304 ttatatagtcatgtttatcattgaacactgaagttctgcgagggctgagcagcact 1363
 Db 73 tttattattcccgtaantttgctcngngatataatgaanaacnngatataangnt 132
 QY 1364 ggcattgaaataataataatcataagtcacaaataaataaataaataa 1416
 Db 133 natgntaaaaataaataaataaataaataaataaataaataaataa 185

RESULT 14
 US-08-924-759-9
 Sequence 9, Application US/08924759
 Patent No. 5962229
 GENERAL INFORMATION:
 APPLICANT: MCGONIGLE, BRIAN
 APPLICANT: O'KEEFE, DANIEL
 TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE
 TITLE OF INVENTION: ENZYMES
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
 STREET: 1007 MARKET STREET
 CITY: WILMINGTON
 STATE: DELAWARE
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 19898
 COMPUTER READABLE FORM:
 MEDIUM TYPE: DISKETTE, 3.50 INCH
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
 SOFTWARE: MICROSOFT WORD VERSION 7.0A
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/924,759
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: FLOYD, LINDA AXAMETHY
 REGISTRATION NUMBER: 33,692
 REFERENCE/DOCKET NUMBER: CL-1128
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 302-892-8112
 TELEFAX: 302-773-0164
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 911 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 TISSUE TYPE: MAIZE
 IMMEDIATE SOURCE:
 CLONE: CC71SE-B.PK0014.B8
 US-08-924-759-9

Query Match 2.7%; Score 38.8; DB 2; Length 911;
 Best Local Similarity 60.4%; Pred. No. 0.062; Indels 0; Gaps 0;
 Matches 64; Conservative 0; Mismatches 42;
 QY 1310 taqcatgttattatcattgaacactgaagttctgcgagggctgagcagcactgcat 1369
 Db 806 TACTCATGTTTTCATCTCTGCAAGTGCACAAACAGTCGCTGCTGTGTTTACTCTTT 865
 QY 1370 gaaataataataataataataataataataataataataataata 1415
 Db 866 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 911

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:55:17 ; Search time 21.41 Seconds
(without alignments)
559.991 Million cell updates/sec

Title: US-09-357-675c-21

Perfect score: 1748
Sequence: 1 MLGFTLRPHRFLSLCPGL.....LPVQHRRLDLYGNLGHPLS 327

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_39.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	519	29.7	307	1 YJM6_YEAST	P47016 saccharomyc
2	490	28.0	272	1 YJM6_SYNY3	P55175 synchocyst
3	448.5	25.7	322	1 YLH8_SCHPO	Q10166 schizosacch
4	420	24.0	291	1 YLH8_YEAST	P49954 saccharomyc
5	356	20.4	262	1 YBEM_ECOLI	P39874 escherichia
6	333	20.2	262	1 YBEM_ECO57	P58054 escherichia
7	262.5	15.0	340	1 Y480_MYCTU	Q11146 mycobacteri
8	254.5	14.6	285	1 YPOQ_PSEFL	P55176 pseudomonas
9	231	13.2	234	1 YAG5_STRAU	P55178 staphylococ
10	221.5	12.7	346	1 NRJ3_ARATH	P46010 arabidopsis
11	220.5	12.6	261	1 YAG5_STRAU	P55177 staphylococ
12	214.5	12.3	346	1 NRJ1_ARATH	P32961 arabidopsis
13	198.5	11.4	339	1 NRJ2_ARATH	P32962 arabidopsis
14	195	11.2	346	1 ALAM_PSEAE	P11436 pseudomonas
15	179.5	10.3	393	1 BUP_RAT	Q03248 rattus norv
16	178.5	10.2	355	1 NRJ4_ARATH	P46011 arabidopsis
17	159	9.1	349	1 NRJ4_TOBAC	Q42965 nicotiana t
18	135	8.9	513	1 YHCK_BACSU	P54608 bacillus su
19	142	8.1	256	1 YAPV_ECOLI	Q47679 escherichia
20	142	8.1	345	1 ALAM_RHOER	Q01360 rhodococcus
21	131	7.5	576	1 NAME2_THEMA	Q9X0Y0 thermotoga
22	115.5	6.6	199	1 YI04_YEAST	P40447 saccharomyc
23	114.5	6.6	679	1 NAME_MYCTU	P71911 mycobacteri
24	113.5	6.5	356	1 NRJ4_ALCEA	P20960 alcyonigenes
25	112.5	6.4	368	1 CYH7_GLOSO	P32964 gliocercos
26	108.5	6.2	552	1 NAME_RHOCA	Q03638 rhodobacter
27	102	5.8	383	1 NRJ1_RHOOR	Q02068 rhodococcus
28	102	5.8	567	1 NAME_AOUAE	Q67091 aquifex aeo
29	100	5.7	365	1 NRJ2_RHOOR	Q03217 rhodococcus
30	98.5	5.6	511	1 LNT_PSEAE	Q92186 pseudomonas
31	97.5	5.6	680	1 NAME_MYCLE	Q9CB66 mycobacteri
32	96	5.5	794	1 SEIL_HUMAN	Q9UBV2 homo sapien
33	94.5	5.4	433	1 FUR4_RAT	Q62994 rattus norv

34	94	5.4	542	1 LNT_CHLTR	O84539 chlamydia t
35	93	5.3	349	1 NRJ6_KLEPO	P10045 klebsiella
36	93	5.3	519	1 LNT_SYNY3	P74055 synchocyst
37	92.5	5.3	520	1 VNN2_HUMAN	Q95498 homo sapien
38	91.5	5.2	541	1 LNT_RHIME	Q52910 rhizobium m
39	91	5.2	512	1 VNN1_MOUSE	Q920K8 mus musculu
40	90	5.1	439	1 LNT_AOUAE	Q67000 aquifex aeo
41	89.5	5.1	1257	1 ERB2_RAT	P06494 rattus norv
42	89.5	5.1	1255	1 ERB2_HUMAN	P04626 homo sapien
43	88	5.0	790	1 SEIL_MOUSE	Q92296 mus musculu
44	88	5.0	2206	1 POLG_POL3L	P03302 poliovirus
45	87.5	5.0	500	1 VNN3_MOUSE	Q9QZ25 mus musculu

ALIGNMENTS

RESULT	1	STANDARD	PRT	307 AA.
YJM6_YEAST				
ID	YJM6_YEAST			
AC	P47016;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	HYPOTHETICAL 34.7 KDA PROTEIN IN SPT10-GCD14 INTERGENIC REGION.			
GN	YJL126W OR J0706.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=5288C / FY1679;			
RX	MEDLINE=97103775; PubMed=8948101;			
RA	Cziedluch C., Kordes E., Pujol A., Jauniaux J.-C.;			
RT	"Sequencing analysis of a 40.2 kb fragment of yeast chromosome X			
RT	reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,			
RT	SPT10, GCD14, RPE1, PBO86, NCA3, ASF1, CCR7, GZF3, two tRNA genes,			
RT	three remnant delta elements and a Ty4 transposon."			
RL	Yeast 12:1471-1474(1996).			
CC	-1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: Z49401; CA899421.1; -.			
DR	SGD; S0003662; YJL126W.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	InterPro: IPR001110; UPF0012.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 307 AA; 34693 MW; 48787CA3B10A828E CRC64;			
Query Match 29.7%; Score 519; DB 1; Length 307;				
Best Local Similarity 37.6%; Pred. No. 6e-39;				
Matches 114; Conservative 56; Mismatches 105; Indels 28; Gaps 6;				
QY	45 ELPLVAVCQVTSPPDKOONFKTCAELVREAARLACACAPLPEAFDFIARPAETLHL---	101		
DB	4 KIKRVANAQGLSSADLTKNLKLVKVELISEALQKADVPFLPEASDYSIQPDLMSRYIAQK	63		
QY	102 SEPLGKGLLEETQVLARECGIWLSG-GFH--ERGD-WEOTQIKYCHVLNSKGAIVA	157		
DB	64 SPFTIRLOASSITDLVVDNSRNIDVSGVHLPESEODLLEGNRVRNVLLIYIDHEGKILO	123		
QY	158 TYRTHLCVDVEIPGQGMCSNSITMPGSLSEPVSTPAGKIGLAVCYDMFPELSLALAQ	217		

```

Db 124 EVOKLLEFVDVDPNCPILKESSKSVOPGKAIPDIIESPLKLSALCYDIRPEPESLKLS 183
QY 218 ACAELITVYSAGSTGPRHMEVLLARAIETOCYVVAAGGRH----- 262
Db 184 MCAELITCFPSATFKIGENHWEILGRARAVDTQCYVLMGQVMDLSPDEMEKOSHMSA 243
QY 263 ---HEKRASGSHMVDPGTVAVARC---SEGPGLCIARIDINLYRLRRLPVFQHRP 316
Db 244 LKSSRSREMGSHMVIDPAGKIIAHNDPSTVGPOLILADLDELLQEIINKKPIYMQRRD 303
QY 317 DLY 319
Db 304 DLF 306

RESULT 2
Y601_SYNY3 STANDARD: PRT; 272 AA.
AC P55175;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 30.2 KDA PROTEIN SLI0601.
GN SLI0601.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kocani H., Sazuka T., Miyajima N.,
RA Sugitara M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 648 to 928 of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D64002; BA10370.1;
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 272 AA; 30191 MW; 7787BA9B2BE003A1 CRC64;

```

```

Query Match 28.0%; Score 490; DB 1; Length 272;
Best Local Similarity 40.2%; Pred. No. 2e-36;
Matches 111; Conservative 46; Mismatches 99; Indels 20; Gaps 7;

```

```

QY 48 LVAAQCVTSPDKQNFCTKAELVREARALGACLAFLPAEDFIADPAETLHLEPLG 106
Db 5 LAAALQMTSPRLNTEMLQAEELIDLAIVQGAELVGLPENFAFLGN---ETEKLEQATAI 61
QY 107 GALLBEYTO-LAREGGLMSTLGGF-----HERGQDMEOQTKIYNCHVILNLSGAVVATPR 160
Db 62 ATATKFKIQTMAQRQVYTLTAGGFPFVAGAG-----KAYNTATLLIAPNGQELARH 114
QY 161 KTHLCQDVEIPGGPMCESNTPGSPLESPPV---STPAKIGLAVCYDMFPELSTALAOA 218
Db 115 KHLHFDVNPDPDNTYWSATVAAQOKY-PPVYHSDSFGMLGISTICYDAVFPPELXRLSLQ 173
QY 219 GAELITVPSAFSGSTGPRHMEVLLARAIETOCYVVAAGGRHHEKRASGSHMVDPW 278

```

```

Db 174 GADVLFVPAFAFYATYKGDHMQVLLQARALENTCYIAPATQCHYERHRTGHAMIDPW 233
QY 279 GTVVAARCSGPGCLCIARIDANLRLRRLPVFQHR 314
Db 234 GYLADAGEKPELAIETAINPDRKOVROOQPSLOHR 269

RESULT 3
YAU8_SCHPO STANDARD: PRT; 322 AA.
ID YAU8_SCHPO
AC Q1016;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 35.7 KDA PROTEIN C26A3.11 IN CHROMOSOME I.
GN SPAC26A3.11.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RA Mclean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z69240; GAA9324.1;
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein.
SQ SEQUENCE 322 AA; 35679 MW; EA6F39B160C7F49F CRC64;

```

```

Query Match 25.7%; Score 448.5; DB 1; Length 322;
Best Local Similarity 32.8%; Pred. No. 1.2e-32;
Matches 108; Conservative 63; Mismatches 123; Indels 35; Gaps 9;

```

```

QY 11 RELSLICPLR--IPLSVLCAQPPRAVAISSSCCLP-----LVAAQCVTSPDKQ 62
Db 4 KFGVLQKTRSFPSLNCYTR---NIMSVASSL-VKDRARIRIGVLOANKKDSSE 59
QY 63 NFKCAELVREARALGACLAFLPAEDFIADP-----AETLHLEPLGKLEBY 113
Db 60 NQLARLKLVEAKNGSNVILPELIFN---SPYGTGFYNOYAEPIEISSP-----SYAL 111
QY 114 TOLARECGMLSLSGFHEHGQDMEOQTKIYNCHVILNLSGAVVATPRTHLCDVEIPGG 173
Db 112 SSMARDTKTYLEGGISIPER---KDGKLYNTAMFDPGKLIARHKLHLEPLDIDPGV 166
QY 174 PGCESNTPGSPLESPPVSTPAKIGLAVCYDMFPELSTALAOGAELITVPSAFSGIT 233
Db 167 SPRESDSLSPGAM-TMWDTEYKFGKLGICYIRPELAMIAARNGCSMITPGAFNLST 225
QY 234 GRAHMEVLLARAIETOCYVVAAGGRHHEKRASGSHMVDPGTVAVARCSGPGCL 293
Db 226 GLHWEILLARARAVNDENMVACCAPARMDADYHSMGSHSTVVDPGKVIATFDEKPSIV 285
QY 294 ARIIDLNLRLRRLPVFQHRPDLXGNL 322
Db 286 ADIDPSVMSSTARNSVPIYTORRPDYSEV 314

```

ID	Y185_YEAST	STANDARD:	PRT:	291 AA.
AC	P49954;			
DT	01-0CT-1996 (Rel. 34, Created)			
DT	01-0CT-1996 (Rel. 34, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	HYPOTHETICAL 32.5 KDA PROTEIN YLR351C.			
GN	YLR351C OR L9638.5.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-S5288C / AB972;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,			
RA	Favelli A., Fulton L., Gattung S., Greco T., Kirsten J.,			
RA	Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,			
RA	Johnson D., Johnston L., Langston Y., Latreille P., Le T.,			
RA	Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,			
RA	Ricken L., Riles L., Taich A., Trevisan E., Vignati D.,			
RA	Wilcox L., Wolhman P., Yaudin M., Wilson R., Waterston R.;			
RL	Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE UPP0012 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/)			
CC	or send an email to license@isb-sib.ch .			
CC	-----			
DR	EMBL: U19102; AAB67751.1; -			
DR	SGD: S0004343; YLR351C.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	InterPro: IPR001110; UPP0012.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPP0012; 1.			
KW	Hypothetical protein.			
SO	SEQUENCE 291 AA; 32549 MW; A813744120088827 CRC64;			
Query Match 24.0%; Score 420; DB 1; Length 291;				
Best Local Similarity 34.5%; Pred. No. 3.6e-30;				
Matches 98; Conservative 53; Mismatches 115; Indels 18; Gaps 7;				
OY	49 VAVQCVT-STDPKQNFRTCAELVDEAR-LGACLAFLPEAFD-----FIARPAET 98			
DB	13 VALVOLSSSPDKMANLDRAATFIERAKNEQPDTKLVLPKECFNSPYSTDQF--RRYSV 70			
OY	LHLSEPLGKLELEETQALRECGWLISGGFERGQDEWQTKIYNCHVLLNSKGAVAT 158			
DB	71 INPKRP--STVQFLSNLANKRKILVGGTIELD---PKTKIYNTSIIFNEDGLIDK 125			
OY	159 YRKTHLCVEIPDGCPMCESNSTMGPSESVSPAGKIGLAVCYDMFPELSIALQA 218			
DB	126 HRKVLFPVDIPNGISPFSESETLSPGEK-STIDIKYGGVGICIDMRFPELAMLARK 184			
OY	219 GAEIITYPSARGSTIGPAHWEVLLFARAIEPCGYVAAAGCRHHEKRSYGHSMVDDP 278			
DB	185 GAFAMITYSARFTVYGPRLHMLLAASRAVDNOYVWALCSPARNLOSSIYAHGHSIVDPR 244			
OY	279 GTVVAACSEGPGLCIARIDLNYLRQLRRHLPVFOHRPDLVGNL 322			
DB	245 GTVAEAGEGEETIYAEIDPEVIESFRQAVPLTKORRFPVSDV 288			

```

ID      YBEM_ECOLI     STANDARD:      PRF:      262 AA.
DT      P39874:P77192:
DT      01-FEB-1995 (Rel. 31, Created)
DT      20-AUG-2001 (Rel. 40, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      HYPOTHETICAL, PROPEIN YBEM.
GN      YBEM OR B0625/B0626.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / MG1655.
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Gishner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.,
RT      "The complete genome sequence of Escherichia coli K-12."
RL      Science 277:1453-1474(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12:
RX      MEDLINE=97061202; PubMed=8905232;
RA      Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA      Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA      Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizubuchi K.,
RA      Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA      Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA      Yano M., Horiiuchi T.,
RT      "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT      corresponding to the 12.7-28.0 min region on the linkage map."
RL      DNA Res. 3:137-155(1996).
RN      [3]
RP      SEQUENCE FROM N.A.
RA      Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA      Federspiel N., Hyman R., Kaiman S., Komp C., Kurd O., Lew H., Lin D.,
RA      Natch A.S., Oelner P., Roberts D., Schramm S., Davis R.W.;
RN      Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RL      [4]
RP      SEQUENCE OF 1-51 FROM N.A.
RC      STRAIN-K12:
RX      MEDLINE=95075315; PubMed=7984109;
RA      Yamanaoka K., Mitani T., Ogura T., Niki H., Hiraga S.;
RT      "Cloning, sequencing, and characterisation of multicopy suppressors
RT      of a mkb mutation in Escherichia coli."
RL      Mol. Microbiol. 13:301-312(1994).
RN      [5]
RP      SEQUENCE OF 31-78 FROM N.A.
RC      STRAIN-K12 / EMG2;
RA      Robison K., O'Keefe T., Church G.M.;
RL      Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN      [6]
RP      CONCEPTUAL TRANSLATION.
RA      Rudd K.E.;
RL      Unpublished observations (JAN-2000).
CC      -1- SIMILARITY: BELONGS TO THE UP00012 FAMILY.
CC      -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC      DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC      POSITION 67.
CC
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL, AE000167; AAC73726.1; ALT_FRAME.
DR      EMBL, AE000167; AAC73727.1; ALT_FRAME.
DR      EMBL, D90703; BAA35268.1; ALT_FRAME.

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; L13334; AAA17975.1; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 234 AA; 26495 MW; C1CBADCA5E1389A7 CRC64;

Query Match 13.2%; Score 231; DB 1; Length 234;
 Best Local Similarity 31.1%; Pred. No. 1.8e-13;
 Matches 61; Conservative 27; Mismatches 64; Indels 44; Gaps 6;

OY 142 IYVCHVLNKGAVVATRTHTCDVEIPGCGMCESNSTMPPSLESPTSTAG----- 196
 DB 64 IFNTAFADIKTGKVINQYDMHLV-----PMLDEPAFLTAGKNVPE 104
 OY 197 -----KIGLAVCYDMRPPELSLALQAGAEILTY-----PSAFGSITGPAMHEVLLRA 244
 DB 105 TFKLSNGVKYTMQICYDLRPELLRYPARSGATITAFVAVQMPFARLN-----HMQVILKA 159
 OY 245 RALETQCYVVAAGGGRHNEKRASY-GHSMVDPWGTIVARCSEGPGLCLARIDLNYLRQ 303
 DB 160 RALENNMYVIGNCGG--YDGKTQYAGSHVAIINPGEIIEIELSEKELVTVIDIDAVEQ 217
 OY 304 LRHRLPVFOHRREDLY 319
 DB 218 ORKAIPFDSLVPHLY 233

RESULT 10
 NR_L3_ARATH STANDARD; PRT; 346 AA.
 AC P46010; O04909;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE NITRILASE 3 (EC 3.5.5.1)
 GN NIT3 OR AT3G44320 OR T10D17_110.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE=94294436; PubMed=8022831;
 RA Bartel B., Fink G.R.;
 RT "Differential regulation of an auxin-producing nitrilase gene family
 RT in Arabidopsis thaliana.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE=98145459; PubMed=9484465;
 RA Hillebrand H., Bartling D., Weller E.W.;
 RT "Structural analysis of the nitr2/nit1/nit3 gene cluster encoding
 RT nitrilases, enzymes catalyzing the terminal activation step in indole-
 RT acetic acid biosynthesis in Arabidopsis thaliana.";
 RL Plant Mol. Biol. 36:89-99(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;

RX MEDLINE=21016720; PubMed=11130713;
 RA Sallanoubat M., Lemcke K., Rieger M., Ansoorge W., Unseid M.,
 RA Farnham B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Deligny M., Boutry M., Griwell L.A., Maché R., Puidomenech P.,
 RA De Simone V., Choise N., Attienave F., Robert C., Brotier P.,
 RA Wuncker P., Catolico L., Weissenbach J., Saurin W., Queller F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wiedelmann R., Kranz H., Voss H., Holland N., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
 RA Nichteil J., Schafte M., Schoen O., Bariges M., Terol J., Clement J.,
 RA Navaro P., Collado C., Perez-Perez A., Oltenevalder B., Duchemin D.,
 RA Cooke R., Laude M., Berger-Lilaou J.-P., Parnelle B., Masuy D.,
 RA de Haan M., Maarse A.C., Alcazar J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argitrou A., Flores M., Liguori R., Vitale D.,
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,
 RA Cressy T.H., Haas B., Maitl R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Millscher J., Sellers P., Gill J.E., Feldblum T.V.,
 RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Nakayama A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:820-822(2000).
 CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
 CC INDOLE-3-ACETIC ACID.
 CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
 CC -1- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U09959; AAA19627.1; -
 DR EMBL; V07648; CAA68936.2; -
 DR EMBL; AL353865; CAB89000.1; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR00132; Nitril_cyn_hydrolase.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS00920; NITRIL_CHT_1; 1.
 DR PROSITE: PS00921; NITRIL_CHT_2; 1.
 KW Hydrolase; Multigene family.
 FT ACT_SITE 186
 FT ACT_SITE 186
 SQ SEQUENCE 346 AA; 38022 MW; 70CFF421547F2B5E CRC64;

Query Match 12.7%; Score 221.5; DB 1; Length 346;
 Best Local Similarity 27.4%; Pred. No. 2.1e-12;
 Matches 86; Conservative 44; Mismatches 113; Indels 71; Gaps 14;

OY 56 STPDKQNFKTCALVREARALGACLAFLPEAF-----DFIARDPAET 98
 DB 41 ATLDKAKKF-----IVEAASGAKLVLPPEAFIGYPPRGRLGAVGVHNEGGRPEFN 94
 OY 99 LHLSE-PLGKGLIEYTLQALRECGWLWSLGGFHENGQDMQOTKIYNCHVLLNSKAVYA 157
 DB 95 YHSAIKVPGPEVERLAEAGKNVHLVMGATEKDG-----YTYICTALFFSPQGF 148
 OY 158 TYRKTHLCQVE--IPGCGMCESNSTMPPSLESPTSTAGKIGLAVCYDMRPPELSLAL 215
 DB 149 KHRKVPSTLERCIWGG-----DGSITP-----YVDPIGKIGAIQWENRPLLYRTAL 198


```

OY 216 AQAAGAILTYPGAFSGITGPAAHWEVLLARAIETOCYVAAQA-CGR----- 261
DB 199 YAKGEITCAPTADSL-----EMQASMIHINVEGCGFVLSAQFCKRRFPHPDYLND 254
OY 262 -----HHEKRSYSHSNVVDWPGVAVARCS--EGPGLCLARIDLNTLRRLPVPQHR-R 314
DB 255 IVDTEKHPVTSVGGSGVSIISPLGKVLGAPNVESEGLVTADLDGLIARAKLYEDVVGHS 314
OY 315 RPDLYGNL---GHP 325
DB 315 KPDIF-NLTVNEHP 327

RESULT 11
YAG5_STAU STANDARD: PRT: 261 AA.
ID YAG5_STAU STANDARD: PRT: 261 AA.
AC P53177;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 29.8 KDA PROTEIN IN AGR OPERON (ONF 5).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_Taxid=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE GAL;
RX MEDLINE=96004766; PubMed=7565609;
RA Novick R.P., Projan S.J., Kornblum J., Ross H.F., Ji G.,
RA Kreiswirth B., Vandenesch F., Moghazeh S.;
RT "The agr P2 operon: an autocatalytic sensory transduction system in
RT Staphylococcus aureus.";
RL Mol. Gen. Genet. 248:446-458(1995).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52543; CAA36779.1;
DR InterPro: IPR003010; CN_hydrolyase.
DR Pfam: PF00795; CN_hydrolyase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein.
SQ SEQUENCE 261 AA: 29826 MW: 23030958AC18F838 CRC64;

Query Match 12.6%; Score 220.5; DB 1; Length 261;
Best Local Similarity 26.1%; Pred. No. 1.8e-12;
Matches 75; Conservative 53; Mismatches 114; Indels 45; Gaps 12;

OY 45 ELPLV-----AVCVSTPDKQONFKTCALVREARLGIACIAPLEAFDIAPDET 98
DB 7 OLPIYFGDSKNETOITOMFEKMN-----AEV-----DVVVLPKWMN-----NGYDL 49
OY 99 LHLSPGLGKLEEXT-----QLAREGIMLSTGPFHRRGDMQOTQIKYCHVLLNSKGV 155
DB 50 EHLNKKADNNLIGQSFSLKHLKEKYKVDIVAGSVSN-----IRNNOIFMTARSVNKSGL 104
OY 156 VATYKRLCLDVEIPGQCMESNSTM--PGPSLESVPSTPAK--IGLAVCYDMRPPEL 213
DB 105 INEYKVLV-----PMLRHEFLTGEVVAEFPQSLSDTYTQGLICYLDRPELLR 156
OY 214 ALAQAAGAILTYPGAFSGITGPAAHWEVLLARAIETOCYVAAQAOCGRHNEKRASTY-GHS 272
DB 157 YPARSGAKIAFYAQM--PMSRLQHWHSLLKARAENNMVIGTNSG--FDGNTGEVAGHS 213

```

```

OY 273 MVVDWPGVAVARCS-EGPGLCLARIDLNTLRRLPVPQHRPDL 319
DB 214 IYINPGLVGLNLSADILITVDLNLNEVEQORENIPVKSIXLDLY 260

RESULT 12
NR1_ARATH STANDARD: PRT: 346 AA.
ID NR1_ARATH STANDARD: PRT: 346 AA.
AC P32861; Q42543; Q04908;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NITRILASE 1 (EC 3.5.5.1).
GN N111 OR AT3G44310 OR T10D17.100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=CV. LANDSBERG ERCTA; TISSUE=Leaf;
RX MEDLINE=92209532; PubMed=1555601;
RA Bartling D., Seedorf M., Mithoefer A., Weller E.W.;
RT "Cloning and expression of an Arabidopsis nitrilase which can convert
RT indole-3-acetonitrile to the plant hormone, indole-3-acetic acid.";
RL Eur. J. Biochem. 205:417-424(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Zhou L., Bartel B., Thornburg R.W.;
RT "Nucleotide sequence of the Arabidopsis thaliana nitrilase 1 gene.";
RL (In) Plant Gene Register PCR95-130.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98145459; PubMed=9484465;
RA Hillebrand H., Bartling D., Weller E.W.;
RT "Structural analysis of the nitr2/nit1/nit3 gene cluster encoding
RT nitrilases, enzymes catalyzing the terminal activation step in indole-
RT acetic acid biosynthesis in Arabidopsis thaliana.";
RL Plant Mol. Biol. 36:89-99(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,
RA Farmanan B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisme N., Artiguenave F., Robert C., Brotier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saunin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Exfle H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
RA Reinhardt J., Schaefer M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Oltmewelder B., Duchemin D.,
RA Cooke R., Lande M., Berger-Liautaud C., Burnelle B., Masny D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Uteback T., Fujii C.T., Shea T.P.,
RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltischer J., Sellers P., Gali J.E., Feldlyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Saito S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsuno M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinozaki S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";

```

RL Nature 408:820-822(2000).
 RP [5]
 RC CHARACTERIZATION.
 RC STRAIN-CV. LANDSBERG ERECTA;
 RX MEDLINE=94286570; PubMed=8016109;
 RA Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;
 RT "Molecular characterization of two cloned nitrilases from *Arabidopsis thaliana*: key enzymes in biosynthesis of the plant hormone indole-3-acetic acid";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
 CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE INDOLE-3-ACETIC ACID.
 CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, BUT AT A VERY LOW LEVEL DURING THE FRUITING STAGE.
 CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X63445; CAA45041.1; -;
 DR EMBL; X38845; AAB05221.1; -;
 DR EMBL; Y07648; CAA68935.2; -;
 DR EMBL; AL353865; CAB88999.1; -;
 DR PIR; S22398; S22398.
 DR InterPro; IPR003010; CN_hydrolase.
 DR InterPro; IPR000132; Nitril_cyn_hydrolase.
 DR Pfam; PF00795; CN_hydrolase; 1.
 DR PROSITE; PS00920; NITRIL_CHT_1; 1.
 DR PROSITE; PS00921; NITRIL_CHT_2; 1.
 KW Hydrolase; Multigene family.
 FT ACT_SITE 186
 FT CONFLICT 312 312 Y->H (IN REF. 2).
 SQ SEQUENCE 346 AA; 38178 MW; 8D4F887CAD1E3C1F CRC64;

Query Match 12.3%; Score 214.5; DB 1; Length 346;
 Best Local Similarity 27.2%; Pred. No. 8.7e-12;
 Matches 84; Conservative 44; Mismatches 104; Indels 77; Gaps 14;

QY 56 STPDQONFKYCAELVRAARLACLAFLPEAF-----DFTAR 93
 DB 41 ATIDKAERY-----IVRAASKGALVLFPEEGTIGYRGRFGGLAVGVHNEGRDERK 94
 QY 94 DPAETLHLSEPLGKLEETYLQALRECGILSLGFGHERGQDMEQTKIYNCHVLLNS-K 152
 DB 95 YHASAIHVPGEVARLAD---VARKNHVLYVMGAIEKEG-----YTLV-CTVLEFSPQ 143
 QY 153 GAVNAVITYKTHLCDE--IPQGGPMCESNTPGSPLESPTSPGKIGLAVCYMRPE 210
 DB 144 GQFLKRLKMLPTSLERCWGG---DGSTIP-----VYDTPIGKLGAAICWENRML 193
 QY 211 LSLALAQAGAEITVPSPFGSITGPAMHEVLRLARAIEOCVVAAO-CGR----- 261
 DB 194 YRTALYANGIELYCAPTADGS---KEWSSMLHAIIEGCGVLSACQFCQKRPDPHD 249
 QY 262 -----HHEKRASYGHSWVDPMGTVVARS-EGPGICLARIIDNYLRQLRRHL-P 309
 DB 250 YLEFTWDYDKRHSIVSGGSIISPLGQVLAGPNEFESGLTADIDLDGIARAKLTPTS 309
 QY 310 VFQARRPDL 318
 DB 310 VGYSRPDV 318

RESULT 13
 NREL2_ARATH
 ID NREL2_ARATH STANDARD: PRT; 339 AA.

AC P32962; Q96505;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NITRILASE 2 (EC 3.5.5.1)
 GN NIT2 OR AT3G44300 OR T10D17-90.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eusteroideae; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. LANDSBERG ERECTA; TISSUE=Leaf;
 RX MEDLINE=94286570; PubMed=8016109;
 RA Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;
 RT "Molecular characterization of two cloned nitrilases from *Arabidopsis thaliana*: key enzymes in biosynthesis of the plant hormone indole-3-acetic acid";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE=94294436; PubMed=8022831;
 RA Bartel B., Fink G.R.;
 RT "Differential regulation of an auxin-producing nitrilase gene family in *Arabidopsis thaliana*";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Zhou L., Bartel B., Thornburg R.W.;
 RT "Nucleotide sequence of a pathogen induced nitrilase gene from *Arabidopsis thaliana*: Nit2";
 RL (in) Plant Gene Register PGR96-006.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unsel M., Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermayer B., Dalseg M., Boutry M., Grivell L.A., Mache R., Pulgdemenech P., De Simone V., Choise N., Artiguenave F., Robert C., Brotier P., Wincker P., Catolico L., Weissbach J., Sautin W., Quetier F., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V., Wummbach E., Dronek H., Erfle H., Jordan N., Bangert S., Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nakamura G., Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B., Conrad A., Hornischer K., Kauer G., Loehmer T.-H., Nordsiek G., Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Clement J., Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D., Cooke R., Laurie M., Berger-Liauro C., Purrelle B., Masny D., de Haan M., Maarse A.C., Alcaraz J.-P., Cortet A., Casacuberta E., Monfort A., Argitlon A., Flores M., Liguori R., Vitale D., Mannhaupt G., Haase D., Schopf H., Rued S., Zaccaria P., Mewes H.-W., Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J., Rooney T.H., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P., Crasey T.H., Haas B., Mafti R., Wu D., Peterson J., Van Aken S., Pal G., Miltischer J., Sellers P., Gill J.E., Feldblum T.V., Pruss D., Lin X., Nieman W.C., Salzberg S.L., White O., Venter J.C., Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E., Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*";
 RL Nature 408:820-822(2000).
 CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE INDOLE-3-ACETIC ACID.
 CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
 CC -1- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.
 CC -1- DEVELOPMENTAL STAGE: BARELY DETECTABLE IN YOUNG ROSETTES, BUT IS

CC STRONGLY EXPRESSED DURING BOLTING, FLOWERING, AND ESPECIALLY FRUIT
CC DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68305; CAA48377.1; -
DR EMBL: U09958; AAB60275.1; -
DR EMBL: U38845; AAB05220.1; -
DR EMBL: AL353865; CAB86998.1; -
DR PIR: S31969; S31969.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR000132; Nitril_cyn_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS00920; NITRIL_CHT_1; 1.
DR PROSITE: PS00921; NITRIL_CHT_2; 1.
KM Hydrolase; Multigene family.
FT ACT_SITE 179 179 BY SIMILARITY.
FT CONFLICT 37 37 E -> G (IN REF. 3).
FT CONFLICT 48 48 S -> T (IN REF. 3).
SQ SEQUENCE 339 AA; 37153 MW; 06CDE76D2DCC2A47 CRC64;

Query Match 11.4%; Score 198.5; DB 1; Length 339;
Best Local Similarity 25.9%; Pred. No. 2.3e-10;
Matches 78; Conservative 47; Mismatches 115; Indels 61; Gaps 12;

QY 59 DKQNFKTCALVREARLACIALPPEAF-----DPIADPAETLHL 101
DB 31 DHPATLEKANKRIVEAASGSELVWPEAFIGYPRFGGLGVHNEGROEFKRYHA 90
QY 102 SE-PIGSKLLEYTOLARCGLSLSGFHERGQDWEQOKIYNCHVILNSKAAVATPR 160
DB 91 SAIKVPGPEVEKLAELAGNNVYLVMGAIKDG-----YTLCTALFSPQOFLGKR 144
QY 161 KTHLGDVE--IPGQPMCSNSTMPSPLESPPSTPAKIGLAVCYDMPEPELSLAQA 218
DB 145 KLMPTSLERCIINGOG-----DGSITP-----YDPTIGKLAICENRNPLYRTALYAK 194
QY 219 GAEILTYPSAFSITGPAHWEVLLARAIETQCYVVAAG-CGR-----HH 263
DB 195 GIELYCAPPADGS---KEMQSMILHAIIEGCGFVLSACQFLRKDFPDHPLYLFTDWYD 250
QY 264 EKR---ASYGSMVVDPMGYVARCS-EGPGICLARIDLNTYROLRRLH-PYFOHRRPD 317
DB 251 DKPDSIVSGSGSVITISPLGQVLAGPNESEGLITADLDGVARAKLYDSVGHYSRPD 310

QY 318 L 318
DB 311 V 311

RESULT 14
ALAM_PSEAE STANDARD; PRT; 346 AA.
AC P11436;
DT 01-OCT-1989 (Rel. 12, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALIPHATIC AMIDASE (EC 3.5.1.4).
GN AMT OR PA3366.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae.
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE.

RC STRAIN-PAC142;
RX MEDLINE-87219101; PubMed-3108029;
RA Ambler R.P., Auffret A.D., Clarke P.H.;
RT "The amino acid sequence of the aliphatic amidase from Pseudomonas
RT aeruginosa.";
RL FEBS Lett. 215:285-290(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-87219102; PubMed-3108030;
RA Brammar W.J., Charles I.G., Matfield M., Liu C.-P., Drew R.E.,
RA Clarke P.H.;
RT "The nucleotide sequence of the amtE gene of Pseudomonas aeruginosa.";
RL FEBS Lett. 215:291-294(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01.
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: ALLOWS THE ORGANISM TO GROW ON SHORT CHAIN-LENGTH
CC ALIPHATIC AMIDES.
CC -1- FUNCTION: ENABLES THE ORGANISM TO USE ACETAMIDE AS BOTH CARBON
CC AND A NITROGEN SOURCE.
CC -1- CATALYTIC ACTIVITY: A MONOCARBOXYLIC ACID AMIDE + H(2)O = A
CC MONOCARBOXYLATE + NH(3).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M27612; AAA25697.1; -
DR EMBL: AE004759; AAG06754.1; -
DR PIR: A26741; A26741.
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
KM Hydrolase; Complete proteome.
FT CONFLICT 40 40 E -> D (IN REF. 1 AND 2).
FT CONFLICT 169 169 G -> P (IN REF. 2).
FT CONFLICT 317 317 E -> D (IN REF. 1 AND 2).
SQ SEQUENCE 346 AA; 38495 MW; E19CEB474EB92B93 CRC64;

Query Match 11.2%; Score 195; DB 1; Length 346;
Best Local Similarity 25.2%; Pred. No. 4.8e-10;
Matches 76; Conservative 43; Mismatches 129; Indels 54; Gaps 11;

QY 41 SSSCELPYAVCOVSTPDKQNFKTCALVREARL-----GACIALPE-A 87
DB 7 SSSNDYVGVAVVNV-----KMPRLHTAAEVLIDNARKIAEMIVGKQGLPGMDLVVFPREYS 61
QY 88 FDIADPAETLHLSEPLGKLEETQLAREGGLWLSLGFHERGQDWD--TQKIYNC 145
DB 62 LOGIMYDPAEMMETAVAIPEETETISRAOKRANV--GVFSLTGRHEHPKRAYNT 118
QY 146 HVLNLSKGAIVATYRKTHLGDVEIPGQPMCSNSTMPSPLESPPSTPAKIGLAVCYD 205
DB 119 LVLLDNGEIVQYRKTI-----PWCPIEGWTFPGQTYVSEBPKMKISLITCDD 168
QY 206 MRPELSLAQAQAEILT-----YPSAFSITGPAHWEVLLARAIETQCYVVAAGC 259
DB 169 GNPPEIWRDCAMKGAELIVRCGQVMYPAKDOY-----MMAKAMAMANNNVAAVANA 221

OY	34	PRAAIISSSCCELPVAAQCTSTPDDKQ---	NRKTAELVREARAGACIAPLEA---	87
		! : : : ! :	! : : : ! :	
Db	69	POIYRVLGVOKRIPL-----	PTSAVPAEQVSALKRKRIETIEIVAAKMGVNIICQDEAMNM	122
OY	88	-FDPIARDPAETLHLSEPLGSKLLEETYO-	LARECGIMLSIGGFHERGODEOTOKIYNC	145
Db	124	PEACETREKLEPTEFEAESAEDEGLTTRFCQKLAKNHNM-	VVISPLTERDRD--	HGGVLMNT 180
OY	146	HYLLNSKAAVAATRKTHLCDEVTIPGOGPMECSNTPMGPLESPV-	STPAGKTGLAVCY	204
Db	181	AVYISNSGLVWGKTRKKN-----	IPRODPEHSYTYMEG-NLGHFVPTQIQGRIVANICY	234
OY	205	DMRPPELSLALQAQAEILTYPSAFSGITGPAHWEVLLRARAIEITQCYVVAADQGR---	261	
Db	235	GRHHPLMLMTSVNGCAELIIFNPDSATIGELSESMPRIEARNMAIINHCFCTALNVRGOEHY	294	
OY	262	-----	HHKRAVSYGHSNVDPWMCVVVARCSGG-	GGCLARIDDLTYRLQ 304
Db	295	PNETFSGDGKKAHHDLGYFYSSSTVVAAPDGSRTGILSRNODGLVLTENLMLCCOI	350	

Query Match	10.3%;	Score 179.5;	DB 1;	Length 393;
Best Local Similarity	26.0%;	Pred. No. 1.3e-08;		
Matches 77; Conservative	41;	Mismatches 139;	Indels 39;	Gaps 11;

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:54:42 ; Search time 57.61 Seconds
(without alignments)
830.236 Million cell updates/sec

Title: US-09-357-675C-21

Perfect score: 1748

Sequence: 1 MGFTRPPRRFLSLCPGL.....LPVFGHRRPDLXGNLGHPLS 327

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1748	100.0	327	4 076091	076091 homo sapien
2	1487	85.1	323	11 088526	088526 mus musculu
3	1484	84.9	323	11 09114	09114 mus musculu
4	1372	78.5	290	11 09DBF2	09DBF2 mus musculu
5	956	54.7	288	13 091860	091860 xenopus lae
6	710.5	40.6	460	5 076464	076464 drosophila
7	702	40.2	440	5 076463	076463 caenorhabdit
8	645.5	36.9	316	10 091E50	091E50 arabidopsis
9	575.5	32.9	276	3 094660	094660 schizosach
10	537.5	30.7	272	1 027839	027839 methanobact
11	497	28.4	282	2 09HVU6	09HVU6 pseudomonas
12	485.5	27.8	276	4 09NOR4	09NOR4 homo sapien
13	479	27.4	270	2 083040	083040 plectonema
14	474.5	27.1	276	11 09JH2	09JH2 mus musculu
15	436.5	25.0	275	2 09KUV4	09KUV4 vidrio chol
16	433.5	24.8	283	2 09A9Y3	09A9Y3 caulobacter
17	406	23.2	318	10 09LYH1	09LYH1 arabidopsis
18	401	22.9	270	2 09JTI3	09JTI3 neisseria m
19	398	22.8	270	2 09KOW5	09KOW5 neisseria m

20	395.5	22.6	189	11 09CTG9	09CTG9 mus musculu
21	389.5	22.3	113	6 029367	029367 sus scrofa
22	387	22.1	283	5 09VHE4	09VHE4 drosophila
23	371	21.2	231	11 09D0S1	09D0S1 mus musculu
24	342.5	19.6	264	2 09XA70	09XA70 streptomyces
25	318	18.2	297	2 09RRF7	09RRF7 deinococcus
26	317	18.1	270	1 09HIW8	09HIW8 thermoplasma
27	312	17.8	257	1 030121	030121 archaeoglob
28	291.5	16.7	300	10 09XG19	09XG19 lycopersico
29	279	16.0	287	1 09Y9L1	09Y9L1 aeropyrum p
30	274.5	15.7	271	2 09KE11	09KE11 bacillus ha
31	273.5	15.6	292	2 0916J8	0916J8 pseudomonas
32	264.5	15.1	295	2 09PA02	09PA02 xylella fas
33	262	15.0	270	2 091241	091241 pseudomonas
34	255.5	14.6	259	2 091664	091664 bacillus su
35	253.5	14.5	282	2 09Z5U5	09Z5U5 zymomonas m
36	252	14.4	298	12 089413	089413 paramecium
37	249.5	14.3	280	2 09AB15	09AB15 streptomyces
38	248	14.2	292	2 09AB15	09AB15 streptomyces
39	246.5	14.1	280	2 069939	069939 streptomyces
40	241	13.8	220	10 09Z0H4	09Z0H4 arabidopsis
41	240	13.7	292	2 025452	025452 helicobacte
42	239	13.7	292	2 09ZL86	09ZL86 helicobacte
43	238.5	13.6	262	2 09L104	09L104 streptomyces
44	233.5	13.4	271	2 09CB47	09CB47 mycobacteri
45	230.5	13.2	294	2 09ZMC7	09ZMC7 helicobacte

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	327 AA.
076091	076091			
AC	076091			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	NITRILASE HOMOLOG 1.			
GN	NIT1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98337986; PubMed=9671749;			
RA	Pekarsky Y., Campigilio M., Siprashvili Z., Druck T., Sedkov Y.,			
RA	Tilib S., Draganescu A., Wermuth P., Rothman J.H., Huebner K.,			
RA	Buchberg A.M., Mazo A., Brenner C., Croce C.M.,			
RT	"Nitrilase and Fhl1 homologs are encoded as fusion proteins in			
RT	Drosophila melanogaster and Caenorhabditis elegans."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).			
DR	EMBL; AF069984; AAC39901.1; -			
DR	EMBL; AF069987; AAC39907.1; -			
DR	InterPro: IPR001110; UPF0012.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
SO	SEQUENCE 327 AA; 35896 MW; 90FTF99D4BA627B1 CRC64;			
Query Match 100.0%; Score 1748; DB 4; Length 327;				
Best Local Similarity 100.0%; Pred. No. 3.2e-150;				
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1 MGFTRPPRRFLSLCPGLPOLSVLCQPPRRMAISSSCCEPIVAVCOVTSPPK 60			
DB	1 MGFTRPPRRFLSLCPGLRIPOLSVLCQPPRRMAISSSCCEPIVAVCOVTSPPK 60			
OY	61 QONFKTCAELVREARLGCALFLPEAFDFIARDPETHLISPLGKLLSEYTLAREC 120			
DB	61 QONFKTCAELVREARLGCALFLPEAFDFIARDPETHLISPLGKLLSEYTLAREC 120			

```
QY 121 GLMISLGGFHERGQDMEQTOXLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
DB 121 GLMISLGGFHERGQDMEQTOXLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
QY 181 TMPGFSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAGAEILTYPSAFGSTITGPAHMEV 240
DB 181 TMPGFSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAGAEILTYPSAFGSTITGPAHMEV 240
QY 241 LIRARAIETOCYVVAACGGHREKASYGHSVMVDPMCTVVARCSEGGGLCLARIDLNY 300
DB 241 LIRARAIETOCYVVAACGGHREKASYGHSVMVDPMCTVVARCSEGGGLCLARIDLNY 300
QY 301 LRQLRRHLPVFQHRRPDLXGNLGHPLS 327
DB 301 LRQLRRHLPVFQHRRPDLXGNLGHPLS 327

RESULT 2
088526 PRELIMINARY; PRT; 323 AA.
ID 088526:
AC 088526:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE HOMOLOG 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=9837986; PubMed=9671749;
RX SEQUENCE FROM N.A.
RA Pekarsky Y., Campigillo M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wernuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Flit homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL; AF069985; AAC40184.1; -.
DR MGD; MGI:1350916; Ntl1.
DR InterPro; IPR003010; CN_hydrolase.
DR InterPro; IPR001110; UPF0012.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA; 35717 MW; 10151CEB151DF2C7 CRC64;
```

```
Query Match 85.1%; Score 1487; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 1.3e-126;
Matches 276; Conservative 27; Mismatches 20; Indels 4; Gaps 2;
```

```
QY 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
DB 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
QY 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
DB 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
QY 61 QONFKTCAELVEARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
DB 61 QONFKTCAELVEARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
QY 121 GLMISLGGFHERGQDMEQTOXLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
DB 121 GLMISLGGFHERGQDMEQTOXLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
QY 181 TMPGFSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAGAEILTYPSAFGSTITGPAHMEV 240
DB 181 TMPGFSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAGAEILTYPSAFGSTITGPAHMEV 240
QY 241 LIRARAIETOCYVVAACGGHREKASYGHSVMVDPMCTVVARCSEGGGLCLARIDLNY 300
DB 241 LIRARAIETOCYVVAACGGHREKASYGHSVMVDPMCTVVARCSEGGGLCLARIDLNY 300
QY 301 LRQLRRHLPVFQHRRPDLXGNLGHPLS 327
DB 301 LRQLRRHLPVFQHRRPDLXGNLGHPLS 327
```

```
QY 301 LRQLRRHLPVFQHRRPDLXGNLGHPLS 327
DB 297 LQOMRQHLPVFQHRRPDLXGSLGHPLS 323
```

```
RESULT 3
Q9R1N4 PRELIMINARY; PRT; 323 AA.
ID Q9R1N4:
AC Q9R1N4:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=9837986; PubMed=9671749;
RX SEQUENCE FROM N.A.
RA Pekarsky Y., Campigillo M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wernuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Flit homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL; AF069988; AAC40185.1; -.
DR MGD; MGI:1350916; Ntl1.
DR InterPro; IPR003010; CN_hydrolase.
DR InterPro; IPR001110; UPF0012.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA; 35705 MW; FBOD7730713665EF CRC64;
```

```
Query Match 84.9%; Score 1484; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 2.5e-126;
Matches 276; Conservative 26; Mismatches 21; Indels 4; Gaps 2;
```

```
QY 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
DB 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
QY 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
DB 1 MGFTTRPHRLSLICGLRIPLQSLVCAOPRRPAMAISSSSCCLPLVAVCOVSTPDK 60
QY 61 QONFKTCAELVEARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
DB 61 QONFKTCAELVEARLACLAFLPEAFDFIARDAETLHSEPLGKLEXYTLARCC 120
QY 121 GLMISLGGFHERGQDMEQTOXLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
DB 121 GLMISLGGFHERGQDMEQTOXLYNCHVLLNSKGAVALYATYRKTHLCDVEIPGQPMCESNS 180
QY 181 TMPGFSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAGAEILTYPSAFGSTITGPAHMEV 240
DB 181 TMPGFSLESVPSTPAKGKIGLAVCYDMRPELSLALAQAGAEILTYPSAFGSTITGPAHMEV 240
QY 241 LIRARAIETOCYVVAACGGHREKASYGHSVMVDPMCTVVARCSEGGGLCLARIDLNY 300
DB 241 LIRARAIETOCYVVAACGGHREKASYGHSVMVDPMCTVVARCSEGGGLCLARIDLNY 300
QY 301 LRQLRRHLPVFQHRRPDLXGNLGHPLS 327
DB 297 LQOMRQHLPVFQHRRPDLXGSLGHPLS 323
```

```
RESULT 4
Q9DBF2 PRELIMINARY; PRT; 290 AA.
ID Q9DBF2:
AC Q9DBF2:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
```


GN NIT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Williams L.,
RA Wyszewski A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK004988; BAB23723.1; -.
DR MGI; MGI:1350916; Nt1.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
SQ SEQUENCE 290 AA; 31886 MW; 34493DBF2170C71 CRC64;

Query Match 78.5%; Score 1372; DB 11; Length 290;
Best Local Similarity 87.5%; Pred. No. 3e-116;
Matches 252; Conservative 22; Mismatches 14; Indels 0; Gaps 0;

QY 40 SSSCELPLVAVCOVSTPDKOONFKTCAELVREARLGCALFLPAEPFIARDPAETL 99
DB 3 SSTSWELPLVAVCOVSTPDKOONFKTCAELVREARLGCALFLPAEPFIARDPAETL 62
QY 100 HLESEPLGKLEBYTOLARECGIMLSLGFHERGQDEQTOKIYNCVHLNSKGAIVAVY 159
DB 63 LLESEPLNGDLGQYSGQLARECGIMLSLGFHERGQDEQNOKIYNCVHLNSKGAIVAVY 122
QY 160 KRTHLCDVETPQGPWCESNSTMPGSPLESVPSTPAKIGLAVCYDMRPPELSLALQAQ 219
DB 123 KRTHLCDVETPQGPWCESNSTMPGSPLESVPSTPAKIGLAVCYDMRPPELSLALQAQ 182
QY 220 AELTYPSAFSGITGPAHMEVLLRARALTEOCYVVAACGGRHNERASVGHSMVDPWG 279
DB 183 AELTYPSAFSGITGPAHMEVLLRARALTEOCYVVAACGGRHNERASVGHSMVDPWG 242
QY 280 TVVARCSEPGICLARIDLNLRLRHLRVQHRPRDLXGSLGPHLS 327
DB 243 TVVARCSEPGICLARIDLNLRLRHLRVQHRPRDLXGSLGPHLS 290

RESULT 5
QY18G0 PRELIMINARY; PRT: 288 AA.
AC QY18G0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NIT1 PROTEIN 1.
GN NIT1.
OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=10959838;
RA Pace H.C., Hodawadekar S.C., Draganesu A., Huang J., Bieganski P.,
RA Pekarsky Y., Croce C.M., Brenner C.;
RT "Crystal structure of the worm Nit1p1 Rosetta Stone protein reveals a
RT Nit tetramer binding two Fhit dimers.";
RL Curr. Biol. 10:907-917(2000).
DR EMBL; AF284575; AAP87104.1; -.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
SQ SEQUENCE 288 AA; 32061 MW; B3673A80609BF042 CRC64;

Query Match 54.7%; Score 956; DB 13; Length 288;
Best Local Similarity 62.7%; Pred. No. 1.3e-78;
Matches 173; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 47 PLVAVCOVSTPDKOONFKTCAELVREARLGCALFLPAEPFIARDPAETLHSEPLG 106
DB 7 PLVAVCOVSTPDKOONFKTCAELVREARLGCALFLPAEPFIARDPAETLHSEPLG 66
QY 107 GKLEBYTOLARECGIMLSLGFHERGQDEQTOKIYNCVHLNSKGAIVAVYRTKTHCD 166
DB 67 GDIQKRTQIARCGIMLSLGFHERGQDEQTOKISNSHVVDNTHIVYRKALHFD 126
QY 167 VEIPGQGPWCESNSTMPGSPLESVPSTPAKIGLAVCYDMRPPELSLALQAQAEILTYP 226
DB 127 VDIQNGVLSRESSSTLPGALIRPITSPACKIGLGYCDLRFPEFSIALALQAQAEILTYP 186
QY 227 SAFGSLTGPANMEVLLRARALTEOCYVVAACGGRHNERASVGHSMVDPWGTVARCS 286
DB 187 SAFGSLTGPANMEVLLRARALTEOCYVVAACGGRHNERASVGHSMVDPWGTVARCS 246
QY 287 EGPGLCLARIDLNLRLRHLRVQHRPRDLXGSL 322
DB 247 EGPGLCLARIDLNLRLRHLRVQHRPRDLXGSL 282

RESULT 6
QY18G0 PRELIMINARY; PRT: 460 AA.
AC QY18G0;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NIT1 PROTEIN.
GN NIT1PRT OR CG7067.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,


```

09LE50
ID 09LE50 PRELIMINARY: PRT: 316 AA.
AC 09LE50:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-JUN-2001 (TReMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE NITRILASE 1 LIKE PROTEIN.
DE AT4G08790.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA McCombie R.W., Spiegel L.A., Huang E.N., Nascimento L.U.,
RA de la Bastide M., Vil D.M., Preston R.R., Matero A., Shah R.,
RA O'Shaughnessy A., Rodriguez M., Shekher M., Schutz K., See L.H.,
RA Swaby I., Habermann K., Dedhia N.N., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL161813; CAB82115.1; -
DR EMBL: AL161512; CAB78004.1; -
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase.1.
SQ SEQUENCE 316 AA; 34807 MW; 1D740F48DF9A202C CRC64;

Query Match 36.9%; Score 645.5; DB 10; Length 316;
Best Local Similarity 44.2%; Pred. No. 1.9e-50;
Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;

QY 33 RPRAMAISSSCSEPL-----VANCQVSTPDKQONFCTCAELVREARLACL 81
DB 12 KPSLFRITLSSQIPLMATVTKYVAAQMTSVNDLTNFTATCSRLVGEALAGAKL 71
QY 82 AFPEAFDFIARDPAEITLHSEPLGKLEETQLARECGLMISLGFFHREGDMEQTOK 141
DB 72 ICFPENSTFYGDKEGESVKTAERLDGPVMRYCSLARDSNIMLSLGSFQERFD-TH 127
QY 142 IYVCHVLLNSKGAVALVYRTHLCDEVEIPQGGPMCESNSTMPGSLSPVSTPAKIGLA 201
DB 128 LCNTHVVIDDAGIMRIDYQKMLHEDVDVPGSSSKYSESTFVPGKTIYS-VDSPVGRGLT 186
QY 202 VCDVMFPELSLALA-QAGAEILTPSAFSGSTGPAHVEVLLRARALETQCYVAAACG 260
DB 187 VCYDLRFPKTYQQLRFEGKQAVLLVPSAFKTYGGEAHMETILRARALETQCYVAAACG 246
QY 261 RHHEKRAVSCHSVWVDPGTVVARCSE--GPGCLARIDINYLQLRRLHPV 310
DB 247 KHHEKESYSTDTLLIDPMGTIVGRLPDVRVSTGIYVADIDSLIDSVTKMPI 298

RESULT 9
ID 094660 PRELIMINARY: PRT: 276 AA.
AC 094660:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HYPONETICAL 30.4 KDA NITRILASE-LIKE PROTEIN C651.02 IN CHROMOSOME II.
GN SPBC651.02.

```

```

OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Volkert G.,
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE N-TERMINAL OF NITRILASES.
DR EMBL: AL035570; CAB37598.1; -
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase.1.
KW Hypothetical protein; Lyase.
FT ACCT_SITE 156 156 BY SIMILARITY.
SQ SEQUENCE 276 AA; 30421 MW; 07AA741A54297E07 CRC64;

Query Match 32.9%; Score 575.5; DB 3; Length 276;
Best Local Similarity 44.2%; Pred. No. 3.4e-44;
Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;

QY 46 LPLVAVCVSTPDKQONFCTCAELVREARLGAFLPAPFPIARDPAETLHL-SEP 104
DB 1 MTLAAVAVQNSSSGSLKNAICAKELISQAAKAKCFFPPASDFIAHNSDEALITNHP 60
QY 105 LGGKLEETQLARECGLMISLGFFHREGDMEQTOKYVCHVLLNSK-----GAVVA 157
DB 61 DCSKFIDVRESAKRHSIFVNI-CVHE-----PSVKKN-KLLNSLFIPLHGEIIS 110
QY 158 TYRTHLCDEVEIPQGGPMCESNSTMPGSLSPVSTPAKIGLAVCYDMRPELSLAAQ 217
DB 111 RYSAHLFDEYKIKNGPTLKESNTTLRGEALIPCKTPTGKVSACIDPIRPEQAIKIRN 170
QY 218 AGAEILTPSAFSGSTGPAHVEVLLRARALETQCYVAAACGHHKRRASGSMVDP 277
DB 171 MGAHIIITPSAFTEKTAHMEVLLRARAALDSQCYVLAIPAAGKHNKRASGSMIVDP 230
QY 278 WGTIVARCSE--GP-GCLARIDINYLQLRRLHPVQHRPDL 319
DB 231 WGTIVAGYSDISSPNGLIFADLDLNDVHRTYIPLL--RRNDLY 273

RESULT 10
ID 027839 PRELIMINARY: PRT: 272 AA.
AC 027839:
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE N-CARBAAMOYL-D-AMINO ACID AMIDOHYDROLASE.
GN MTH1811.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=86037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Haritison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
RA Spadafora R., Viccare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000934; AAB86277.1; -
DR InterPro: IPR003010; CN_hydrolase.

```


RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004129; AAF93594.1; -.
DR TIGR; VC0421; -.
DR InterPro; IPR003010; CN_hydrolase.
DR InterPro; IPR001110; UPF0012.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
KW Complete proteome.
SQ SEQUENCE 275 AA; 30630 MM; F1342BBCBF519EE4 CRC64;

Query Match 25.0%; Score 436.5; DB 2; Length 275;

Best Local Similarity 38.1%; Pred. No. 1.3e-31;

Matches 104; Conservative 41; Mismatches 111; Indels 17; Gaps 5;

```
QY 49 VAVCQVSTSPDKQONFKTCALVREARLGAFLPEAFDFIA-----RDPATLHLS 103
| : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 4 VGLVQMTSGPEISRLTLTKQEVAKLAQGAQWIVTPENALLGNREQYHQAESLD--- 60
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 104 PLGGKLEERYQLARECGIWLISLGGFHERGQDWEQOTOKIYNCHVLLNSKGAVAATYRKTH 163
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 61 --HGPOHALASLAKEYGWLIGSMPIRHAEGVTTSSL-----LWNAQGERVAVYDKLH 113
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 164 LCDVET-DGQGMCSNSGMPGSPLESVPSTPAKIGLAVCYDMRPEPELSLALAQAGAEI 222
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 114 MEDVDVADGHQRYRESEFTPEQOV-VVTSPEGALGLSTCYDVRPHLYADLRQGAOI 172
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 223 LTPSAFGSITGPAHMEVLLRAALETQCYVAAACGRHHEKRAASYGSHNVVDPGTVV 282
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 173 LLVPAFTAVTGAHMEVLLRAALETQCYVAAACGRHHEKRAASYGSHNVVDPGTVV 232
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 263 ARCSGPGICLARIPLNLRLRLHLPVFQHR 315
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 233 ANLGAEVQSKVVEFDLATLDSVRRAMPITQHR 265
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: April 26, 2002, 17:02:22
Job time: 460 sec

A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A:Reference number: A85001; MUID:20083488

A:Accession: D85088

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <STO>

A:Cross-references: GB:NC_001268; NID:g7267521; PIDN:CAB78004.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4G08790

A:Map position: 4

Query Match 36.9%; Score 645.5; DB 2; Length 316;

Best Local Similarity 44.2%; Pred. No. 1.3e-45;

Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;

33 RPRMAISSSCCLP-----VAVCVSTPDKQONFKTCAELVREARLGLAFLPAFPDIADPAETLHL-SEP 104

12 KPSFTRTITSSQIPLTWATTNKTVRVAAQMTSVNDLMTNFATCSKRLVDEAALAGAKL 71

82 AFLPEAFDPIARDPAETLHLSEPLGKLEBYTOLARECGMLSLGCFHREGDWEQTOK 141

72 ICFPENSFVADKESGSKIAEPLDGPWERYCSLARSNTLWLSLGFQERPD-----TH 127

142 IYNCHVLNSKAVVATYRTKTHLCDVEIPGQPMCESNSTMPGSLSPVSTPAKIGLA 201

128 LCNTHVVIDAGMIRDTYQKMHLEFDVDPGSSYKESFTVPGTKIVS-VDSPGRLGLT 186

202 VCYDMRPPELSLALA-QAGAEILTPPSAFSGSITGPAHVEVLLRAAITEQCYVAAACG 260

187 VCYDLRFKRIYQOLRFEGKACQVLLVPSAFVTGTGEAHWEILLRAAITEQCYVAAACG 246

261 RHHEKRSYGHSMVVDPMGTAVARCS--GPGICLARTDLYLRQLRRHLV 310

247 KHNEKREYGGTLIIDPMGTIVVGRLPDRVSTGIYADIDFSLIDSVKTKMPI 298

RESULT 3

49 VAVCVSTPDKQONFKTCAELVREARLGLAFLPAFPDIADPAETLHLSEPLGK 108

3 LGICOMQYTERKKNVGTAVEMIVAEARGARLIVLPMF-----TCPYVYN 49

109 LLEEYTYQ-----LARECGMLSLGCFHREGDWEQTOKLYNCHVLNSKAVV 156

50 LFEYAEEDENCYSTRTMSIARELGILHVAQSIP-----EPREGIYNTSVMDGGMII 104

157 ATYKTHLCDVEIPGQPMCESNSTMPGSLSPVSTPAKIGLAICYDMRPELSLALA 216

105 GKRRVHLFDINVGELISFRESDSLADSV-TVETPQCYWGVGICVDMRPELSRMA 163

217 QAGAEILTPPSAFSGSITGPAHVEVLLRAAITEQCYVAAACGHHHEKRSYGHSMVVD 276

164 LGCAEVLFPGFAMTWTTPAHMRLLVRSKALDNCYCYAASPARPSAIVAYGHSMLAD 223

277 PMGTAVARCSGPGICLARTDLYLRQLRRHLVPOHRRPDLVG 320

224 PMGSVVCADASPVTADIDLDEAVERIRRLPLLRNRPDVG 267

Query Match 32.9%; Score 575.5; DB 2; Length 276;

Best Local Similarity 44.2%; Pred. No. 1.7e-43;

Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;

45 LPLVAVCVSTPDKQONFKTCAELVREARLGLAFLPAFPDIADPAETLHL-SEP 104

1 MTLAAVAVCVSTPDKQONFKTCAELVREARLGLAFLPAFPDIADPAETLHL-SEP 104

105 LGKLLLEBYTOLARECGMLSLGCFHREGDWEQTOKLYNCHVLNSK-----GAVVA 157

61 DCSKFIKRVRSATKHSIFVNI-CVHE-----PSKRN--KLINSSLFIEPLHGLIIS 110

158 TYRTHLCLDVEIPGQPMCESNSTMPGSLSPVSTPAKIGLAICYDMRPELSLALAQ 217

Db 111 RYKRAHLFDEIVEIKNGPTLKESNTTLRGEAILPCKTPLGKVSAGICEDIRPEQAIRLN 170

218 AGAEILTPPSAFSGSITGPAHVEVLLRAAITEQCYVAAACGHHHEKRSYGHSMVVD 277

171 MGAHITTPSAFTERTGAHWEVLLRAALDSQCTVIAAPAGCKHNEKRSYGHSMVVD 230

278 MGTAVARCS--GP-GICLARTDLYLRQLRRHLVPOHRRPDLVG 319

231 MGTAVARCS--GP-GICLARTDLYLRQLRRHLVPOHRRPDLVG 319

231 MGTAVARCS--GP-GICLARTDLYLRQLRRHLVPOHRRPDLVG 319

RESULT 4

B69109

N-carbamoyl-D-amino acid amidohydrolase - Methanobacterium thermoautotrophicum (strat

C:Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999

C:Accession: B69109

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.

; Qiu, D.; Spadafora, R.; Viscare, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani,

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MUID:98037514

A:Accession: B69109

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-272 <MTH>

A:Cross-references: GB:AE000934; GB:AE000666; NID:92622924; PIDN:AAB86277.1; PID:9262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1811

C:Superfamily: hypothetical protein YLR351c

Query Match 30.7%; Score 537.5; DB 2; Length 272;

Best Local Similarity 38.4%; Pred. No. 4e-40;

Matches 109; Conservative 49; Mismatches 95; Indels 31; Gaps 4;

49 VAVCVSTPDKQONFKTCAELVREARLGLAFLPAFPDIADPAETLHLSEPLGK 108

3 LGICOMQYTERKKNVGTAVEMIVAEARGARLIVLPMF-----TCPYVYN 49

109 LLEEYTYQ-----LARECGMLSLGCFHREGDWEQTOKLYNCHVLNSKAVV 156

50 LFEYAEEDENCYSTRTMSIARELGILHVAQSIP-----EPREGIYNTSVMDGGMII 104

157 ATYKTHLCDVEIPGQPMCESNSTMPGSLSPVSTPAKIGLAICYDMRPELSLALA 216

105 GKRRVHLFDINVGELISFRESDSLADSV-TVETPQCYWGVGICVDMRPELSRMA 163

217 QAGAEILTPPSAFSGSITGPAHVEVLLRAAITEQCYVAAACGHHHEKRSYGHSMVVD 276

164 LGCAEVLFPGFAMTWTTPAHMRLLVRSKALDNCYCYAASPARPSAIVAYGHSMLAD 223

277 PMGTAVARCSGPGICLARTDLYLRQLRRHLVPOHRRPDLVG 320

224 PMGSVVCADASPVTADIDLDEAVERIRRLPLLRNRPDVG 267

RESULT 5

S56907

hypothetical protein YJL126w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein J0706

C:Species: *Saccharomyces cerevisiae*

C>Date: 05-May-1995 #sequence_revision 08-Sep-1995 #text_change 29-Sep-1999

C:Accession: S56907

R:Chaplin, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.

submitted to the Protein Sequence Database, September 1995

A:Reference number: S56907

A:Accession: S56907

A:Molecule type: DNA

A:Residues: 1-307 <C21>

A:Cross-references: EMBL:Z49401; NID:g1008323; PIDN:CAA89421.1; PID:g1008324; MIPS:YJ

A: Experimental source: cultivar Columbia; BAC clone F14F18
C: Genetics:
A: Map position: 5
A: Introns: 25/3; 64/3; 139/3; 182/1; 207/1; 233/2; 238/3; 265/3; 293/2
A: Note: F14F18.210

Query Match	23.2%;	Score 406;	DB 2;	Length 318;
Best Local Similarity	33.7%;	Pred. No. 2.2e-28;		
Matches 102;	Conservative 45;	Mismatches 116;	Indels 40;	Gaps 7;

```

OY 46 LPL-----VAVCOVTPSPDQCOHEKTCACELVYREAALGCLAFPLPAFPEADRA 96
Db 16 LPLAPLPTKFNICQLQSLVSTQDKKRNISNAKKAIIEPAASKGAKVLVLPPTMNSPYSNDS 75
OY 97 ETLHLEPPLGKLLVEYTQOLARECGWLST---GGEHERGQDMEOTOKLYNCHVLLNSK 152
Db 76 FPVYAEELDAGGDASPSTAMLSIEVSKRLKLTIIIGGSIPEBYGVD---RLYNCCVFGSD 130
OY 153 GAVATYRKTKHLCVELEIPGCGPMCESNSTMPGPELES-----VSTPAKGITG 199
Db 131 GELAKAKHKHILHPDIDIPKRTFMEKSTIPLAG---EFTPLVDGYNGNLGPLENIIPDGRIG 187
OY 200 LAVCTDMRPEPLSLALQACAEIITYSAFGSIITGPAHMEVLLRARAIEFGCYVVAADOC 259
Db 188 IGCIYDIRFOELAIYARAGNHILCYGCAENMTGTPRLHWEILORARATDQIYV---ATC 244
OY 260 GRHHEKRAST---GHSWVDMWGTVVAVRCSBGPLCLARLDVNLRYRHLNLFVQHRP 316
Db 245 SPARDSGAGYAMWHSILVGRFGEVLATTHEHEAIIIAELDIYSLEGRKISLPLNQRRC 304
OY 317 DLY 319
Db 305 DLY 307

```

RESULT 12
E81834
conserved hypothetical protein NMA2044 [imported] - Neisseria meningitidis (strain Z2491)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: E81834
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
: Holroyd, S.; Jajels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556
A:Accession: E81834
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-270 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85262.1; PID:g738067
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: opaB; NMA2044
C:Superfamily: hypothetical protein YUR351c

	Query Match	22.9%	Score 401;	DB 2;	Length 270;
	Best Local Similarity	37.3%	Pred. No. 4.9e-26;		
	Matches 103;	Conservative 35;	Mismatches 112;	Indels 26;	Gaps 8;
Qy	49 VAVCVSTSPKQONFKTCAELVREARLGLACFLPEAPFRIARDPATLHLSEPL-CG	107			
	: : : : : : : : : : : : : : : : : : : :				
Db	6 VAAQVMGSGVSPENVAAMKRLVARAAGVDWVLLPEYWLMGANDIDKTLALAEPLGG	65			
Qy	108 KLEELYTLARECGIMLSLG- ----FHERQDMEQTKIYNCHVLYLNSKGAVAATYRKT	162			
	: : : : : : : : : : : : : : : : : : : :				
Db	66 RFQTLSESTAKECGVVL-FGGTVPLQSPDEAG- ----KVNNTLLVYGRDGKRTGLYHKM	117			
Qy	163 HLCDEIVEGGQPMCSNSNTMPGSP- ----LESPYSPACKIGLACYCDMRPFLSLALAA	218			
	: : : : : : : : : : : : : : : : : : : :				
Db	118 HLPFGSGGLGE-RYAADPTLGLREVPHLLAGGVSAAG- ----LCYDVRPEE-FERRQL	169			

[illegible]

RESULT 13
B81199
nitrilase NMB041 [imported] - Neisseria meningitidis (strain MC58 serogroup B)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: B81199
R:Reteljin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Li, H.; Qin, H.; Yamathavan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: B81199
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-270 <TEF>
A:Cross-references: GB:AE002400; GB:AE002098; NID:g7225659; PIDN:AAF40879.1; PID:g7222222
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB041
C:Superfamily: hypothetical protein YLR351c

Query Match	22.8%	Score	398	DB	2	Length	270
Best Local Similarity	36.8%	Prod. No.	9.1e-28				
Matches	100	Conservative	33	Mismatches	121	Indels	18
				Gaps			6
QY	49	VAVCVSTPDKQONFKTCALVELREARLGIACLAFLPEAFDPIARDPAETLHSEPL	CG	107			
Db	6	VAVVMVSGVSEPTVNAAMKRLVARAEAGMVLVLPKTVMLGANDITKLALAEPLGG	65				
QY	108	KLLEETOLARECGIWLISLG-----FHBDGDWDEQTKIYNCHVLYNSKGAIVATRYKT	162				
Db	66	RQGTALSEFAKCGVYL-FGTVPLDQSCAG-----KVMNTLVYGGDGYRTGTYHKM	117				
QY	163	HLCDVEIPGGPMCESNSTMPGSPLESFVSTPAGKTIGLAIVCYDMRPPELSLAAQGAEL	222				
Db	118	HL-FGFSLSIGRYEADTIRAGGDVPHLSAGVPAVPAAGICVDYRPE--FFRRQLPFDV	173				
QY	223	LTVPAFGSITTPAHEWVILRARALEITOCYVAAAOCCGHHHEKRASYSMSMVDPGTV	282				
Db	174	LMLPAAFHTTTCKAHMEILLRARAVENTCCYVAAAOCCGHHHEKRASYSMSMVDPGTV	233				
QY	283	ARCSEGPGLCLARIDLNTLRQLRHLPVQHR	314				
Db	234	DVLPBEGEGVYADIDANRLNSVNNRLPALAKYR	265				

RESULT 14
t36488
probable hydrolase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: t36488
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21608
A:Accession: t36488
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-264 <SNV>
A:Cross-references: EMBL:AL096822; PIDD:CA846930.1; GSPDB:GN00070; SCODEB:SCGD3.08C

THIS PAGE BLANK (USPTO)

DR WPI: 2000-171195/15.
DR N-PSDB: AAZ46101.
XX
PT Novel nitrilase homologs used as diagnostic and therapeutic reagents
PT for the detection and treatment of cancer -
XX
XX Disclosure; Fig 6; 25pp; English.
XX
XX The present sequence is encoded by the coding region of human, murine,
CC Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC human and mouse NIT1 genes are members of an uncharacterised
CC mammalian gene family with homology to bacterial and plant nitrilases.
CC The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC for fusion proteins in which the Fhit domain is fused with a Nit domain.
CC In mouse and humans, FHIT and NIT are encoded by two different genes,
CC localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC chromosomal fragile site FRA3B, is often altered in most common forms
CC of human cancer. The NIT1 protein overcomes the mutated inactivation
CC of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC and analogues of them, and antibodies are used as diagnostic and
CC therapeutic reagents for the detection and treatment of cancers.
XX
XX
S0 Sequence 464 AA:

Query Match 99.7%; Score 1743; DB 21; Length 464;
Best Local Similarity 100.0%; Pred. No. 2e-174;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LGFTTRPHRFLSLCPGLRIPOLSVLCAPPRAMAISSSSCEPLVAVCVSTPDKO 61
DB 37 lgtftrphrflslcpqlrtpqlsvlcqpramaissscelplvavcvstpdtkq 96
QY 62 QNEFKCAELVREARLGAFLPEAFDPIARDPAETLHLSEPLGKLEETQLAREG 121
DB 97 qnfekcaelvreaarlgaflpeafdiardpaetlhlseplgklleeytqlaregg 156
QY 122 LMLSIGFHERGQDWEOTQKIYNCHVLNSKGAVVATYRKTHLCVETPGCPMCESNST 181
DB 157 lmlsigfhergqdweqtqkiynchnvlnskgavvatyrkthlcvetpggpmcesnst 216
QY 162 MPGPLESFVSTPACKIGLAVCYDMRPPLSLALQAGAEILTYBSAGSITGPAHWEVL 241
DB 217 mpgplesfvstpacgkiglavcydmrpfelslalaqageelltybsagstgpahtwevl 276
QY 242 LRAARLENGCYVVAAGCGRHHKERSYGHSMVDPWGVVAVARCEGPELCIARDIANL 301
DB 277 lraarleqcyvvaagcgrhhekrasyghsmvdpwgvavarcsepglclardianlyl 336
QY 302 ROLRRHLPEVFOHRRPDLVGNLGPPLS 327
DB 337 rqlrrhlpevfqhrppdlygnlghpls 362

RESULT 2
AAB57054
ID AAB57054 standard; Protein: 224 AA.
XX
AC AAB57054;
XX
XX 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1632.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytoskeletal; cardioprotective; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotoxic; anti-infective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
XX Homo sapiens.
OS

XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05988.
PF
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
XX WPI: 2000-587513/75.
DR N-PSDB: AAF16557.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
XX
PS Claim 11; Page 2090-2091; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytoskeletal,
CC cardioprotective, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotoxic, anti-infective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 224 AA:

Query Match 53.0%; Score 927; DB 21; Length 224;
Best Local Similarity 94.7%; Pred. No. 4.1e-89;
Matches 179; Conservative 3; Mismatches 3; Indels 4; Gaps 3;

QY 1 MGFTTRPHRFLSLCPGLRIPOLSVLCAPPRAMAISSSSCEPLVAVCVSTPDKO 60
DB 26 mgftrphrflslcpqlrtpqlsvlcqpramaissscelplvavcvstpdtk 85
QY 61 QNFKTCAELVREARLGAFLPEAFDPIARDPAETLHLSEPLGKLEETQLAREG 120
DB 86 qnfktcaelvreaarlgaflpeafdiardpaetlhlseplgklleeytqlarec 145
QY 121 GLMLSIGFHERGQDWEOTQKIYNCHVLNSKGAVVATYRKTHLCVETPGCPMCESNST 180
DB 146 glmlsigfhergqdweqtqkiynchnvlnskgavvatyrkthlcvetpgp99-1c-vka 203
QY 181 TWP--GPSL 187
DB 204 tlpcjlpvl 212

RESULT 3
AAV10877
ID AAV10877 standard; Protein: 153 AA.
XX
XX AAV10877;
AC
XX
XX 14-MAY-1999 (first entry)
DT
XX Amino acid sequence of a human secreted protein.
DE
XX
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
XX

XX developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..46
FT /note= "signal peptide"
FT Protein 47..153
FT /note= "secreted protein"
XX
XX WO9907891-A1.
XX
XX 18-FEB-1999.
PD
XX
XX 04-AUG-1998; 98WO-US16235.
XX
XX 19-AUG-1997; 97US-0056732.
PR 05-AUG-1997; 97US-0054798.
PR 05-AUG-1997; 97US-0054803.
PR 05-AUG-1997; 97US-0054804.
PR 05-AUG-1997; 97US-0054806.
PR 05-AUG-1997; 97US-0054807.
PR 05-AUG-1997; 97US-0054808.
PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055316.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056365.
PR 19-AUG-1997; 97US-0056366.
PR 19-AUG-1997; 97US-0056557.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
XX
XX WPI: 1999-167452/14.
DR N-PSDB: AAX30398.
XX
XX New isolated human genes encoding secreted polypeptides - useful for
FT diagnosis and treatment of pathological diseases
FT
XX
XX Claim 3; Page 304; 331pp; English.
PS
XX The specification describes secreted proteins and their corresponding
CC polynucleotides which are useful for preventing, treating or ameliorating
CC medical conditions, e.g. by protein or gene therapy. Pathological
CC conditions can also be diagnosed by determining the amount of the
CC secreted polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the products, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, cardiovascular disorders,
CC prostate diseases, asthma, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.

XX
SQ Sequence 153 AA:
Query Match 47.0%; Score 821; DB 20; Length 153;
Best Local Similarity 100.0%; Pred. No. 3,3e-78;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 KCESNSTMGPSPLESFVSTPAGKIGLAVCYDMRFPPLSLALAOAGELITTPSAFSTTG 234
DB 1 mcesnstmpsplesfvsstpagkiglavcydmrfpplsllalagaelittpsaftstg 60
QY 235 PAHWEVLRLARAIENQCYVAAAGCGRHHKRSASVGHSMVDPWGVVARSCEGPGICLA 294
DB 61 pahwevlrlaratetqcyvvaagcgrhnekrsasvghsmvdpwgtvvarcsegpgicla 120
QY 295 RIDLNTYLRQLRRLHLPVQHRRPDLVGNLCHPLS 327
DB 121 ridlntyrlrqlrrhlpvfqhrtrpdlvgnlchpls 153
RESULT 4
AAB80984
ID AAB80984 standard; Protein; 276 AA.
XX
XX AAB80984;
AC
XX 12-JUN-2001 (first entry)
DT
XX
XX Human hnit3-ase.
DE
XX
XX Human hnit3-ase; Cushing's adrenal gland tumour;
KW nitrile hydrolytic enzyme; nitrilase.
XX
XX Homo sapiens.
OS
XX CN127398-A.
PN
XX 27-DEC-2000.
PD
XX
XX 30-MAY-2000; 2000CN-0116221.
PF
XX
XX 30-MAY-2000; 2000CN-0116221.
PR
XX
XX (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.
XX
XX Xu X, Qian B, Zhang X;
PI
XX
XX WPI: 2001-245678/26.
DR N-PSDB: AAF84214.
XX
XX Human nitrile hydrolytic enzyme protein and its coding sequence -
PT
XX
XX Claim 2; Page 11; 20pp; Chinese.
PS
XX
XX The present invention relates to human hnit3-ase (nitrilase) protein,
CC which is expressed in Cushing's adrenal gland tumour, and its coding
CC sequence (AAF84214 and AAB80984). The present invention also relates to a
CC preparation method of said protein and nucleic acid sequence, and a
CC method of detecting human hnit3-ase nucleic acid sequence and polypeptide
CC in sample.
XX
SQ Sequence 276 AA:
Query Match 27.8%; Score 485.5; DB 22; Length 276;
Best Local Similarity 36.7%; Pred. No. 1.4e-42;
Matches 105; Conservative 46; Mismatches 98; Indels 37; Gaps 6;
QY 49 VAVCQYTPDRQONFKTCALVREARLQACIALPFAFDIARDPAETLHLEPLGSK 108
DB 6 laligqssiksdnvttracsfiraatgaktivspecf-----nspygak 52

```

QY 109 LLEEF-----IQLAREGJLMLSLGFIHREGDMQOTKIRYNCHVLSKNAVY 156
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 53 yfpeaekipgestbqlsevakoesilylgsispe----edagkllyntcavfgrpdtll 107
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 157 AYRRKTHLCADVEIRPOGCMCESNSTMPSPLESPESTRPAGKIGLAVCYDMRFPPELSALA 216
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 108 akyrthihfdidvphkitfgesktlspdsf-stfdtdyrcvvglygicdmfiaeaaqly 166
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 217 QNGAEILTYPSAFSGITSPAHMEVLLRLARAIEOCYVVAAMAAOCGRHNEKRASY--GHSM 273
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 167 gqgcqdlvyppgfnltitgpbawellqrsravdngvyataspa---rdckesyawgshst 223
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 274 VVDPRGYTVARCSSEPGCLARIDNLNTYRLQRHLPVQRHRRDLY 319
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 224 vvnpgewlakaqteaalvysdidlklkaelrqglpvrfrqkrdsly 269
    ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 5
AAB42967
ID AAB42967 standard; Protein: 159 AA.
AC AAB42967;
XX
XX
DT 08-FEB-2001 (first entry)
XX
XX Human OREFX ORF2731 polypeptide sequence SEQ ID NO:5462.
DE
XX
XX Human; open reading frame; OREFX; detection: cytostatic; hepatotropic;
KW humaner; antiparotatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopothic; antiarthritic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
XX
XX WO200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 31-MAR-2000; 2000WO-US086621.
PE
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CUPRA-) CUPRAGEN CORP.
PA
XX
XX Shimkets RA, Leach M;
PI
XX
XX WPI: 2000-602362/57.
DR N-PSDB; AAC77176.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4643; 5507pp; English.
PS
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human OREFX open reading frames 1 to 3161. The OREFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antiparotatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC

```

[illegible]

PT -
XX Claim 17; SEQ ID NO: 6568; 246bp + Sequence Listing; English.
PS
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
CC
XX
SQ Sequence 266 AA;

Query Match 16.2%; Score 283.5; DB 22; Length 266;
Best Local Similarity 32.3%; Pred. No. 2.3e-21;
Matches 90; Conservative 41; Mismatches 115; Indels 33; Gaps 10;

OY 49 VAVCGVSTPPKQKNCALVREARLGAFLPEAFDPIARDPATLHL---SEPL 105
DB 3 IALLQISTNSDMDKMDFAILLRDAEAKAEGGARVLVFPFA---TSQSFQIGRIDTQAEEL 58
OY 106 GSKLELEYTOLARECGMLSLGFFH---ERGDDECTOKIYNCHVLNSGAVAVATPR 160
DB 59 dgeftavrkhladelidvllvlgmftpradvtgq---ektisvntvllsagsg-lhgqyn 114
OY 161 KTHLCDVEIPCGGRCESNSTMPGPSLESVSTPRAKIGLAVCYMRPELSLAAQAQA 220
DB 115 khtyda-----fygreesdtkvpgdel-vvfevddkikfgvacydtkrfpegfkdlarnga 168
OY 221 ELTVPSAFGSITGPA---HMEVLLRARALETQCVCVVAAGCGRHNERK-----ASYGH 271
DB 169 qllvprtswgqd-gpqrklegwvllprataldstcwlvacggarlpeelrderktgptqlygh 226
OY 272 SMVYDPMGTVVAVARCSEPGGLCLARIDLNYLRRLRPV 310
DB 227 smvtnphgvevlasagyepeemladdidvsglaklrealpr 265

RESULT 7
AAG09683 ID AAG09683 standard; Protein: 299 AA.
XX
AC AAG09683;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7710.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
OS
XX
XX EPI033405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141844.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148968.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

```
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 15.4%; Score 270; DB 21; Length 299;
Best local Similarity 27.6%; Pred. No. 7.3e-20;
Matches 85; Conservative 47; Mismatches 124; Indels 52; Gaps 10;

```
QY 48 LVAVQVSTPDKOONFTCAELVREARLGLACLAFLPEAF-----DFTARDBA 96
DB 11 vvsslqfacsddistvaaraerliveaakganilllqelllegyyfcgaqredffkrakp 70
QY 97 ETLHLSEPLGLKLEEVYQLARECGILMLSLGFERGODWEQTKIYNCHVLNLSKGAVV 156
DB 71 yknh-----ptlarmkklakeigvnpvsffe-----eantkynsialiddgtdl 117
QY 157 ATRYKTHLCDEVEIPGOGPMCESNSTM-DPSLESVPSTPAGKIGLAVCYDMKFPPLSLAL 215
DB 118 gtyrkshpdp-----gpyggekfyfnpdgtcfkvyfqtckfakigvpaicwdyffpaaram 171
QY 216 AQAGAEILTYVSAPGSG-----ITGPAHWEVLLRARATETQCYVVAACGCR-----HH 263
DB 172 vlgaeellfypaigsepdqglldsrldwrrvmgghaganvvpvlasnlgkellsetehg 231
QY 264 EKRRAS-YGHSWVVDPMGTGVVAVRCSE-GPGLCLARIDLNVYLRQLRHLPLVFGHRRPDLV-- 319
DB 232 psqitfygtsfslagptgeivaeadkksaevlvaqfdlmlkikrsgwyvfrdrrpdljykv 291
QY 320 -----GNTL 322
DB 292 llmldgnl 299
```

RESULT 9

```
AAB79700
ID AAB79700 standard; Protein; 252 AA.
XX
AC AAB79700;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:134.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-1B00923.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 14-JUL-1999; 99US-0148613.
PR 12-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
```

```

PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.

XX
XX (BAD1 ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX MPI: 2001-137957/14.
XX
XX N-PSDB; AAF71819.
XX
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
XX PT pathway proteins, useful for producing fine chemicals in
XX PT microorganisms, including organic acids, nonproteinogenic amino acids,
XX PT and purine and pyrimidine bases -
XX
XX Claim 20, Page 361-362; 1737pp; English.
XX
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
XX CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
XX CC MP nucleic acids are useful for the production of fine chemicals
XX CC in microorganisms, including organic acids, nonproteinogenic amino
XX CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
XX CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
XX CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX
XX Sequence 252 AA:

Query Match 14.7%; Score 256.5; DB 22; Length 252;
Best Local Similarity 32.5%; Pred. No. 1.5e-18;
Matches 86; Conservative 36; Mismatches 110; Indels 33; Gaps 10;

QY 63 NECTCAELVEAARLGLCLFLPEAFPIARDPAETLHL---SEPLGKLEETQLARE 119
DB 3 nfallidaaeakaaggarvivfpea---tsgsfgtgrltgdeelgetstavrklade 58
QY 120 CGLWLSLGFH-----ERGDWEGTOKIVNCHVLLNSKGVAVATYRKHLCDDVIEIPQGP 174
DB 59 ldvivvaagmtfpadvtrvg---ektlsrvnvtvlisag-ihgynkhiyda-----fg 109
QY 175 MCSNSTMPGSLSPSTPAGKIGLAVCYDMRPPELSLQAQAEILYPSAFGSITG 234
DB 110 yresdvtkpgdel-vvfevddikftgvatcydirfpegkdlnargagqilivpitswgd--g 166
QY 235 PA--HMEVLLRALETQCYVVAQAACGRRHNEK-----ASTGSHMNVYDPMGTVVARC 285
DB 167 pgklegewipraraldstcwiavagqarlpeelrderktpgtgishmvpnphgevlasa 226
QY 286 SEGPGICLARIIDLYLRLRRLHPV 310
DB 227 gyepemliadivsglakirealp 251

RESULT 10
ID AAB80193
ID AAB80193 standard; Protein; 252 AA.
XX
XX AAB80193;
XX
XX 30-APR-2001 (first entry)
XX
XX Corynebacterium glutamicum MP protein sequence SEQ ID NO:1120.
XX
XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
XX KW fine chemical production; microorganism; organic acid; nucleoside;
XX KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
XX KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

```

```

KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
XX
XX Corynebacterium glutamicum.
XX
XX WO200100843-A2.
XX
XX 04-JAN-2001.
XX
XX
XX 23-JUN-2000; 2000WO-IB00923.
XX
XX 25-JUN-1999; 99US-0141031.
XX PR 01-JUL-1999; 99DE-1030476.
XX PR 02-JUL-1999; 99US-0142101.
XX PR 08-JUL-1999; 99DE-1031415.
XX PR 08-JUL-1999; 99DE-1031418.
XX PR 08-JUL-1999; 99DE-1031419.
XX PR 08-JUL-1999; 99DE-1031420.
XX PR 08-JUL-1999; 99DE-1031424.
XX PR 08-JUL-1999; 99DE-1031428.
XX PR 08-JUL-1999; 99DE-1031434.
XX PR 08-JUL-1999; 99DE-1031435.
XX PR 08-JUL-1999; 99DE-1031443.
XX PR 08-JUL-1999; 99DE-1031453.
XX PR 08-JUL-1999; 99DE-1031457.
XX PR 08-JUL-1999; 99DE-1031478.
XX PR 08-JUL-1999; 99DE-1031510.
XX PR 08-JUL-1999; 99DE-1031541.
XX PR 08-JUL-1999; 99DE-1031573.
XX PR 08-JUL-1999; 99DE-1031592.
XX PR 08-JUL-1999; 99DE-1031632.
XX PR 08-JUL-1999; 99DE-1031634.
XX PR 08-JUL-1999; 99DE-1031636.
XX PR 09-JUL-1999; 99DE-1032125.
XX PR 09-JUL-1999; 99DE-1032126.
XX PR 09-JUL-1999; 99DE-1032130.
XX PR 09-JUL-1999; 99DE-1032186.
XX PR 09-JUL-1999; 99DE-1032206.
XX PR 09-JUL-1999; 99DE-1032227.
XX PR 09-JUL-1999; 99DE-1032228.
XX PR 09-JUL-1999; 99DE-1032229.
XX PR 09-JUL-1999; 99DE-1032230.
XX PR 14-JUL-1999; 99DE-1032922.
XX PR 14-JUL-1999; 99DE-1032926.
XX PR 14-JUL-1999; 99DE-1032928.
XX PR 14-JUL-1999; 99DE-1033004.
XX PR 14-JUL-1999; 99DE-1033005.
XX PR 14-JUL-1999; 99DE-1033006.
XX PR 12-AUG-1999; 99US-0148613.
XX PR 27-AUG-1999; 99DE-1040764.
XX PR 27-AUG-1999; 99DE-1040765.
XX PR 27-AUG-1999; 99DE-1040766.
XX PR 27-AUG-1999; 99DE-1040832.
XX PR 31-AUG-1999; 99DE-1041378.
XX PR 31-AUG-1999; 99DE-1041379.
XX PR 31-AUG-1999; 99DE-1041380.
XX PR 31-AUG-1999; 99DE-1041384.
XX PR 31-AUG-1999; 99DE-1041396.
XX PR 03-SEP-1999; 99DE-1042076.
XX PR 03-SEP-1999; 99DE-1042077.
XX PR 03-SEP-1999; 99DE-1042079.
XX PR 03-SEP-1999; 99DE-1042086.
XX PR 03-SEP-1999; 99DE-1042087.
XX PR 03-SEP-1999; 99DE-1042088.
XX PR 03-SEP-1999; 99DE-1042095.
XX PR 03-SEP-1999; 99DE-1042124.
XX PR 03-SEP-1999; 99DE-1042129.
XX PR 09-MAR-2000; 2000US-0187970.
XX
XX (BAD1 ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX
XX

```

DR WPI: 2001-137957/14.
DR N-PSDB: AAF72312.
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PR pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX Claim 20; Page 1678-1679; 1737pp; English.
PS
XX
CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polypeptides and enzymes.
CC
XX
SQ Sequence 252 AA:

Query Match 14.7%; Score 256.5; DB 22; Length 252;
Best Local Similarity 32.5%; Pred. No. 1.5e-18;
Matches 86; Conservative 36; Mismatches 110; Indels 33; Gaps 10;

OY 63 NFKTCAELVREARLACACLAFLPEAFDLIARDPAETLH--SEPLGKILEEYTLARE 119
DB 3 nfallrdaekaaegargylvfrea---tsqsfygrldtqgeeldgefstavrkade 58
OY 120 CGLMLSLGFIH---ERQDWEQOKITNCHVNLNSKCAVATYTKTILCYEITRGQP 174
DB 59 ldlvvaagmflpbadlvqrg---ektlsrvnnctvlisgaq-lhgygnkhlthya----fg 109
OY 175 MCESNSTMGPSPLESPSPACKIGLAVCYDMRFPPLSLALAOAGAEITLYPSAFGSITG 234
DB 110 yresdvtvkgpdel-vvfeyddikfygatydirfpeqfdlarnsqdlivpwtsgd--g 166
OY 235 PA---HWEVLLRARALETCQYVAAACGRHHEKR-----ASYGSMVVDPEWGVVARC 285
DB 167 pklqevwlvpraladstcswlvacqgarlpeelrderkxprtgigsmvtnpbgviala 226
OY 286 SEPGICLARIDLVYRQRLRPV 310
DB 227 gyepemladiavsglakirealp 251

RESULT 11
AAG09684
ID AAG09684 standard; Protein; 220 AA.
XX
AC AAG09684;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7711.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135639.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161927.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 13.8%; Score 241; DB 21; Length 220;
 Best Local Similarity 29.6%; Pred. No. 5.2e-17;
 Matches 68; Conservative 41; Mismatches 87; Indels 34; Gaps 8;

OY 115 OLARCCGMLSLGCFHERGQMEQOTKIYNCHVLLNSKGAVVATYRKTHLCDEVEIPGQGP 174
 Db 3 klakelgyvviyrstfe-----eantahynsaiidadtldiglyrkshlpd-----gp 50
 OY 175 MCESTSM-PPGSLSPVSTPAKIGIAVCYDMRPELISLALAQGAELITYPSPAFGS-- 231
 Db 51 gyqekfyfnpdclgfkvfqtkfakigvalcwgqwfpeaarawlqgaellfypalqsep 110
 OY 232 ---ITGPAHWEVLRLARAIEFQCYVVAACQGR-----HHEKRAS-YGHSMVVDPMCT 280
 Db 111 qdgdidsrdhwrvmgnganvplvasnrlgkclsetehpsqiltfygtsfiagptge 170
 OY 261 VVARCSE-GPGICLARIDLNTLRQLRHLPVQHRPDIY-----GNL 322
 Db 171 lvaeadksaevlvaqfdldmikskrgswgvtirtrpdiylkvllmdgnl 220

RESULT 13
 ID AAG81841 standard; Protein; 264 AA.

AC AAG81841;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:776.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis;

KW vaccination; endocarditis.

OS Staphylococcus epidermidis.

MO200134809-A2.

PN XX

PD 17-MAY-2001.
 PF 09-NOV-2000; 2000MO-US30782.
 PR 09-NOV-1999; 99US-0164258.
 PA (GLAX) GLAXO GROUP LTD.
 PI kimmerly wj;
 DR WPI; 2001-316495/33.
 DR N-PSDB; AAH52691.
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 useful for vaccinating against infections, e.g. endocarditis -
 PS Claim 18; Page 239; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the S. epidermidis polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention. N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

Sequence 264 AA;

Query Match 12.5%; Score 218; DB 22; Length 264;
 Best Local Similarity 31.2%; Pred. No. 1.8e-14;
 Matches 58; Conservative 35; Mismatches 69; Indels 24; Gaps 7;

OY 142 IYNCHVLLNSKGAVVATYRKTHLCDEVEIPGQGP-----CRSNSTMPGSLSPVSTP 194
 Db 94 lynthatahknknelneykvhlv-----pmlrepfldcgn-vvpepfylsdqtl- 143
 OY 195 AGRIGIAVCYDMRPELISLALAQGAELITYPSPAFSITGPAHWEVLRLARAIEFQCYV 254
 Db 144 ---ltqilcydlrffellfyparkgakiayvaqws-erldhwslilkaratendilfiv 199
 OY 255 AAAQGRHHEKRASY-GHSMVVDPMCTVVARCSEGGICLARIDLNTLRQLRHLPVQHR 313
 Db 200 acnscg--ddghtnyagnsivlnpgeiildhldckegviltlthvdvldqgryeyivtrn 257
 OY 314 RRPDIY 319
 Db 258 lkpnlly 263

RESULT 14
 ID AAG44266 standard; Protein; 340 AA.

AC AAG44266;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55427.

PN XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PE 25-FEB-1999; 9905-0121825.
PX 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126254.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0131449.
PR 04-MAY-1999; 9905-0132048.
PR 04-MAY-1999; 9905-0132407.
PR 05-MAY-1999; 9905-0132484.
PR 06-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0132863.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 18-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139452.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.

PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141099.
PR 01-JUL-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 02-JUL-1999; 9905-0142154.
PR 06-JUL-1999; 9905-0142055.
PR 08-JUL-1999; 9905-0142390.
PR 09-JUL-1999; 9905-0142803.
PR 12-JUL-1999; 9905-0142920.
PR 13-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 16-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144332.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147008.
PR 04-AUG-1999; 9905-0147203.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147132.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149358.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.

PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	13-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154179.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157573.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159328.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	

Query Match 12.4%; Score 217.5; DB 21; Length 340;

Best Local Similarity 26.4%; Pred. No. 2.9e-14;

Matches 87; Conservative 47; Mismatches 128; Indels 67; Gaps 14;

OY	41	SSSCELPLVAVCOVSTPDKOONFTGALVREARLGAFLPRAF-----	88
DB	16	sstvtvtlvqsttyn--dtpatlhkaekffiveaaskgkylvlfpaafiggypgrffigl	73
OY	89	-----DFLARDGAFLHLSE-PLGCKLLEFYQLARECGLMISLGCFHERGQDMEOTOKI	142
DB	74	avgvlnheegrdefrnyhasaalkvppxverlaelagknvhlvmgaiekdg-----yl	127
OY	143	YNCHVLNLSKGAVVATYRTHLCDVE--IPGOGPMCESNSTMPGPLESPSTPAKIGL	200
DB	128	yctalfspgqgflgkhrkvmptstlercivwgq-----dgsflp-----vydtpikiga	177
OY	201	AVCYDMRPEELSLAQAQAEILTYPSAFGSITGPAHWEVLRLARAIEMQCYVVAQAQ-C	259
DB	178	alcwenmplytatalyakieileycaptadysl-----ewqasmihavegscfvlshqfc	233
OY	260	GR-----HHEKRSYGHSMVNDPWGTVNAACS-EGEGCLARIDLNT	300
DB	234	krrefrepdyllndivdtkendprvtvsgsgsvllspdgkvlagprnyeseglvadidldg	293
OY	301	LRQLRRHLVFOH-RRPDLYGNL---GHP	325
		: : : : : : : :	

DB	294	iarakiyfdvghyskpdfi-nltvnehp	321
		RESULT 15	
		AAG44265	
ID		AAG44265 standard; Protein; 346 AA.	
XX			
AC		AAG44265;	
XX			
DT		18-OCT-2000 (first entry)	
XX			
DE		Arabidopsis thaliana protein fragment SEQ ID NO: 55426.	
XX			
KW		Protein identification; signal transduction pathway; metabolic pathway;	
KW		hybridisation assay; genetic mapping; gene expression control; promoter;	
KW		termination sequence.	
XX			
OS		Arabidopsis thaliana.	
PN		EP1033405-A2.	
XX			
PD		06-SEP-2000.	
XX			
PF		25-FEB-2000; 2000EP-0301439.	
XX			
PR		25-FEB-1999;	99US-0121825.
PR		05-MAR-1999;	99US-0123180.
PR		09-MAR-1999;	99US-0123548.
PR		23-MAR-1999;	99US-0125788.
PR		25-MAR-1999;	99US-0126264.
PR		29-MAR-1999;	99US-0126785.
PR		01-APR-1999;	99US-0127462.
PR		06-APR-1999;	99US-0128234.
PR		08-APR-1999;	99US-0128714.
PR		16-APR-1999;	99US-0129845.
PR		19-APR-1999;	99US-0130077.
PR		21-APR-1999;	99US-0130449.
PR		23-APR-1999;	99US-0130510.
PR		23-APR-1999;	99US-0130891.
PR		28-APR-1999;	99US-0131449.
PR		30-APR-1999;	99US-0132048.
PR		30-APR-1999;	99US-0132407.
PR		04-MAY-1999;	99US-0132484.
PR		05-MAY-1999;	99US-0132485.
PR		06-MAY-1999;	99US-0132486.
PR		06-MAY-1999;	99US-0132487.
PR		07-MAY-1999;	99US-0132483.
PR		11-MAY-1999;	99US-0134256.
PR		14-MAY-1999;	99US-0134218.
PR		14-MAY-1999;	99US-0134219.
PR		14-MAY-1999;	99US-0134221.
PR		14-MAY-1999;	99US-0134370.
PR		18-MAY-1999;	99US-0134768.
PR		19-MAY-1999;	99US-0134941.
PR		20-MAY-1999;	99US-0135124.
PR		21-MAY-1999;	99US-0135353.
PR		24-MAY-1999;	99US-0135629.
PR		25-MAY-1999;	99US-0136021.
PR		27-MAY-1999;	99US-0136392.
PR		28-MAY-1999;	99US-0136782.
PR		01-JUN-1999;	99US-0137222.
PR		03-JUN-1999;	99US-0137528.
PR		04-JUN-1999;	99US-0137502.
PR		07-JUN-1999;	99US-0137724.
PR		08-JUN-1999;	99US-0138094.
PR		10-JUN-1999;	99US-0138540.
PR		10-JUN-1999;	99US-0138847.
PR		14-JUN-1999;	99US-0139119.
PR		16-JUN-1999;	99US-0139452.
PR		16-JUN-1999;	99US-0139453.
PR		17-JUN-1999;	99US-0139492.
PR		18-JUN-1999;	99US-0139454.
PR		18-JUN-1999;	99US-0139455.


```

Db      134 yctalfispqgflgkhrkwmptslercldwgq-----dgstlp-----vydtpjgkiga 183
OY      201 AVCYDMRFPPELSIALAQAAGAEILTFPSAFCSITGPAHMEVLARAAIETOCYVVAQAQ-C 259
Db      184 alcwenmplyrtalyakglelcapladysl---ewqasmihlavegscfvlshqfc 239
OY      260 GR-----HHEKRSYSGHSMVVDPMGTFVVARCS-EGPGLCLARIDLNY 300
Db      240 krrelfepndylnfdiavtkkehdpvtvggsvlispjgkvlagpnysesglvtadidldgd 299
OY      301 LRQLRRRLPFVQH-RRPDLYGNL---GHP 325
Db      300 iarakiyfdvvgbhyskpdif-nltvnehp 327

```

Search completed: April 26, 2002, 16:54:36
 Job time: 234 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:50:57 ; Search time 28.88 Seconds
(without alignments)
254.798 Million cell updates/sec

Title: US-09-357-675C-21
Perfect score: 1748
Sequence: 1 MLGFTTRPHRFLSLCPGL.....LPVQHRRLDLYGNLGHPLS 327

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/pdata/2/1aa/3A_COMB.pep:*
2: /cgn2_6/pdata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/pdata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/pdata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/pdata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/pdata/2/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.5	11.5	303	2	US-08-294-871A-64 Sequence 64, Appl
2	200.5	11.5	303	3	US-08-876-398A-64 Sequence 64, Appl
3	199.5	11.4	303	2	US-08-294-871A-26 Sequence 26, Appl
4	199.5	11.4	303	2	US-08-294-871A-40 Sequence 40, Appl
5	199.5	11.4	303	2	US-08-294-871A-44 Sequence 44, Appl
6	199.5	11.4	303	3	US-08-876-398A-26 Sequence 26, Appl
7	199.5	11.4	303	3	US-08-876-398A-40 Sequence 40, Appl
8	199.5	11.4	303	3	US-08-876-398A-44 Sequence 44, Appl
9	198.5	11.4	303	2	US-08-294-871A-6
10	198.5	11.4	303	2	US-08-294-871A-34 Sequence 34, Appl
11	198.5	11.4	303	2	US-08-294-871A-60 Sequence 60, Appl
12	198.5	11.4	303	3	US-08-876-398A-6 Sequence 6, Appl
13	198.5	11.4	303	3	US-08-876-398A-34 Sequence 34, Appl
14	198.5	11.4	303	3	US-08-876-398A-62 Sequence 62, Appl
15	197.5	11.3	303	3	US-08-294-871A-62 Sequence 62, Appl
16	197.5	11.3	303	3	US-08-876-398A-62 Sequence 62, Appl
17	196.5	11.2	303	2	US-08-294-871A-10 Sequence 10, Appl
18	196.5	11.2	303	2	US-08-294-871A-12 Sequence 12, Appl
19	196.5	11.2	303	2	US-08-294-871A-22 Sequence 22, Appl
20	196.5	11.2	303	2	US-08-294-871A-38 Sequence 38, Appl
21	196.5	11.2	303	2	US-08-294-871A-42 Sequence 42, Appl
22	196.5	11.2	303	2	US-08-294-871A-46 Sequence 46, Appl
23	196.5	11.2	303	2	US-08-294-871A-54 Sequence 54, Appl
24	196.5	11.2	303	2	US-08-294-871A-58 Sequence 58, Appl
25	196.5	11.2	303	2	US-08-294-871A-68 Sequence 68, Appl
26	196.5	11.2	303	3	US-08-876-398A-10 Sequence 10, Appl
27	196.5	11.2	303	3	US-08-876-398A-12 Sequence 12, Appl

28 196.5 11.2 303 3 US-08-876-398A-22 Sequence 22, Appl
29 196.5 11.2 303 3 US-08-876-398A-38 Sequence 38, Appl
30 196.5 11.2 303 3 US-08-876-398A-42 Sequence 42, Appl
31 196.5 11.2 303 3 US-08-876-398A-46 Sequence 46, Appl
32 196.5 11.2 303 3 US-08-876-398A-54 Sequence 54, Appl
33 196.5 11.2 303 3 US-08-876-398A-58 Sequence 58, Appl
34 196.5 11.2 303 3 US-08-876-398A-68 Sequence 68, Appl
35 195.5 11.2 303 1 US-07-917-111-2 Sequence 2, Appl
36 195.5 11.2 303 1 US-07-917-111-3 Sequence 3, Appl
37 195.5 11.2 303 1 US-08-479-638-2 Sequence 2, Appl
38 195.5 11.2 303 1 US-08-479-638-3 Sequence 3, Appl
39 195.5 11.2 303 2 US-08-294-871A-2 Sequence 2, Appl
40 195.5 11.2 303 2 US-08-294-871A-4 Sequence 4, Appl
41 195.5 11.2 303 2 US-08-294-871A-8 Sequence 8, Appl
42 195.5 11.2 303 2 US-08-294-871A-14 Sequence 14, Appl
43 195.5 11.2 303 2 US-08-294-871A-16 Sequence 16, Appl
44 195.5 11.2 303 2 US-08-294-871A-18 Sequence 18, Appl
45 195.5 11.2 303 2 US-08-294-871A-20 Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-294-871A-64
Sequence 64, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
NUMBER OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294, 871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971, 758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917, 111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211, 641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990

FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: 904136
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-64

Query Match 11.5%; Score 200.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred. No. 7.2e-14;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 46 LVAVQO---VSTPRKQONFKTCALVREARLACLAFLPE--AFDFLAR----DPAET 98
DB 5 ILAVGQGGPIARARETREQVAVRLDLMTKAASGANFIVPELALTFPPRMHFTDEAL 64
QY 99 LHLSE-PLGSKLLEETQALAREGGLMSLGGFH---ERGQDWEQTOIKYNCHVLLNSKA 154
DB 65 DSFYETEMPGPVYRPLFEKAALGIGFNLGYALVEGG-----VKRRNTSLVDKSK 119
QY 155 VVATYRKTHLDCVEIPGQGPCESTNMPGSPLESVPSTP-----AGKIGLAVC 203
DB 120 IVGKTRKTHL-----PGHK---EYEA YRPFQHLKRYFEPGDLGFPVYVDVDAKMGMFIC 171
QY 204 YDMRPELSLALAOAGAEILTYPSAFGSTTGPAMHEVL-----LRARAIET 249
DB 172 NDRRPEAMRWVGLRGAEILIC-----GGYNTPTNHPDIPVQHDLTSFNHLLSMQAGSYON 226
QY 250 QCYVVAACGGRHHEKRAYSYGMVVDPMGTVVARSCE-GPGICLARIDLNTYRQRLRH 308
DB 227 GAMSAAAGKVGW-EENCMLLGHSCIVAPRGETVALTTLEDEVITRAAVDLDCRELREHI 285
QY 309 PVF-OHRPDLVG 320
DB 286 FNFQKHQROPHYG 298

RESULT 2
US-08-876-398A-64
Sequence 64, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NAKABA, Hirokazu
APPLICANT: TAKANO, Masayuki

APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-876-398A-64

Query Match 11.5%; Score 200.5; DB 3; Length 303;
Best Local Similarity 26.2%; Pred. No. 7.2e-14;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 48 LVAVQO---VSTPRKQONFKTCALVREARLACLAFLPE--AFDFLAR----DPAET 98
DB 5 ILAVGQGGPIARARETREQVAVRLDLMTKAASGANFIVPELALTFPPRMHFTDEAL 64
QY 99 LHLSE-PLGSKLLEETQALAREGGLMSLGGFH---ERGQDWEQTOIKYNCHVLLNSKA 154
DB 65 DSFYETEMPGPVYRPLFEKAALGIGFNLGYALVEGG-----VKRRNTSLVDKSK 119
QY 155 VVATYRKTHLDCVEIPGQGPCESTNMPGSPLESVPSTP-----AGKIGLAVC 203
DB 120 IVGKTRKTHL-----PGHK---EYEA YRPFQHLKRYFEPGDLGFPVYVDVDAKMGMFIC 171
QY 204 YDMRPELSLALAOAGAEILTYPSAFGSTTGPAMHEVL-----LRARAIET 249
DB 172 NDRRPEAMRWVGLRGAEILIC-----GGYNTPTNHPDIPVQHDLTSFNHLLSMQAGSYON 226
QY 250 QCYVVAACGGRHHEKRAYSYGMVVDPMGTVVARSCE-GPGICLARIDLNTYRQRLRH 308
DB 227 GAMSAAAGKVGW-EENCMLLGHSCIVAPRGETVALTTLEDEVITRAAVDLDCRELREHI 285

OY 309 PVF-OHRRDLXG 320
Db 286 ENFKOHPOHYG 298

RESULT 3

US-08-294-871A-26

Sequence 26, Application US/08294871A

Patent No. 5824522

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro

APPLICANT: Namba, Hirokazu

APPLICANT: Takano, Masayuki

APPLICANT: Yajima, Kazuyoshi

APPLICANT: Yamada, Yukio

APPLICANT: Takahashi, Satomi

APPLICANT: Okubo, Kazuma

APPLICANT: Yamada, Kazuhiko

APPLICANT: Hiraiishi, Yoshiro

TITLE OF INVENTION: Immobilized Enzyme Preparation and

TITLE OF INVENTION: Process for Producing D-a-Amino Acid

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294,871A

FILING DATE: 22-AUG-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/971,758

FILING DATE: 12-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917,111

FILING DATE: 07-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/211,641

FILING DATE: 11-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 400848/1990

FILING DATE: 07-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP91/01696

FILING DATE: 06-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 407922/1990

FILING DATE: 27-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 078840/1991

FILING DATE: 11-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 140051/1991

FILING DATE: 12-JUN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP92/00739

FILING DATE: 10-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 212692/1992

FILING DATE: 10-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 340078/1992

FILING DATE: 21-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wegner, Harold C.

REGISTRATION NUMBER: 25,258

REFERENCE/DOCKET NUMBER: 74129/127/AOPA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 303 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-294-871A-26

Query Match

Best Local Similarity 26.98; Score 199.5; DB 2; Length 303; Matches 83; Conservative 47; Mismatches 129; Indels 49; Gaps 14;

OY 48 LVAVCO---VTSPPDKQNEFKTCAELVREARLGCALFLPE--AFDFIAR----DPAET 98
Db 5 ILAVGQGGPRAARETREQVYVRLDMITRAASGANFIYFPELALITFFPRMHTDEAEL 64
OY 99 LHLSE-PLGKLELEYTOLARECGLWLSIGF--ERGQDWEQOTKINYCHVLLNSKA 154
Db 65 DSFEYTEMGPVYRPLFEKAELGIFNLGYALVYEGG----VKRRFTSLVYKSGK 119
OY 155 VVATYRKTLCVDEIFGCGPMCSNSTMPGSLSPYSTP-----AGTGLAVC 203
Db 120 IYGVYRKHIL-----PGHK---EYEAYPEQHLKRYFEGDGLGFYVDVDAKMGAFIC 171
OY 204 YDRPPELSLALAQAGAEIT--YPSAFGSIQPAHMEVL-----LRARALETCYV 254
Db 172 NDRPPEAMRWMLRGAEIICGGYNPTNPVPOHDHLISPHHLSMCGSGYONGAMSA 231
OY 255 AAACGRHHEKRSYGHSMVDPWGTIVVARCSE--GGLCLARIDLVNLRRLRLPYF-Q 312
Db 232 AAGKVGK--EENCMLLGHSCIVAPTGEIVALTTLLEDEVITAANDLDRCRLRHETFNKQ 290
OY 313 HRRPDLXG 320
Db 291 HROPQHYG 298

RESULT 4
US-08-294-871A-40
Sequence 40, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4079222/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 0788440/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-40

	Query Match	Similarity	Best Local	Conservative	Matches	Score	DB 2	length
	11.4%	26.2%	16.2%	46%	126%	199.5	303	
OY	48	LVAVQ----	YTPPRKQNKETCAELVRAAALGACIALPPE--AFDFIAR----	DPAET	98			
					
DB	5	ILVAGQGPFIARAEPIREQVVRVRLMLTKTAASRGANFIYEPFELALTFPPRHMFIDEAL		64				
					
OY	99	LHLSE-PLGKILLEYTOLARECGMLSLGGRH---	EQGDWEQOTQIKYNCHVLLNSKGA	154				
					
DB	65	DSFVETMEGQPVAVRLFEKAAALGIGFNGVAVLVEGG----	VKRRTNLSLVDKSKR	119				
					

```

OY      204 YDMRPFELSLAALGAAGELITTPSAFGSTGPAHWEVL-----LRARAIET 249
Db      172 NDRMRPEAMRVMGRLGAEIIC-----GGYMRPTNPPLVPQHDILTFEHHLSMQASGYCN 226

OY      250 QCYYVAAAACGCRHNEHKRASTGSHSNVDPKMGTVVARKCE--GRGLCLALIDINLYRQLRRHL 308
Db      227 GAWSAAGAAGKACM--EENCMLLGHSCIVAPGTEIYALLTTLEDEVIYTAAVDLDRCRELRHT 285

OY      309 PVF--QHRRPDLYG 320
Db      286 FNFQKHROPQHTG 298

```

RESULT 5
US-08-294-871A-44
Sequence 44, Application US/08294871A
Patent No. 5824522

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanda, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991

;; FILING DATE: 12-JUN-1991
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP92/00739
;; FILING DATE: 10-JUN-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: JP 212692/1992
;; FILING DATE: 10-AUG-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP93/01101
;; FILING DATE: 05-AUG-1993
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: JP 340078/1992
;; FILING DATE: 21-DEC-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wegner, Harold C.
;; REGISTRATION NUMBER: 25,258
;; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 672-5300
;; TELEFAX: (202) 672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 44:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 303 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-294-871A-44

Query Match 11.4%; Score 199.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred. No. 9.3e-14;

Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 48 LVAVCQ---VSTPRKQONFKTCALVREARLACLAFLPE--AFDFIAR---DPAET 98
DB 5 ILAVGGGPIARAEETREQVVRLLDMITKASRGANFIVPELATTFEPPRWHTDEAEL 64
QY 99 LHLSE-PLGGKLEETQALRECGMLSLGFGH---ERGQDWEQOKIYNCHVLLNSKA 154
DB 65 DSFEYTEMGPVVRPLPEKKAELGIGFNILGYAELVVEGG-----VKRFNFSILVDKSGK 119
QY 155 VVATYRRTKHLCDVEIPGGPWCESNSMPGSLSPVSTP-----AGKIGLAVC 203
DB 120 IYGYRKRIHL-----PGHK---EYEAYRPFQHLEKRYEPGDLGFPYVDVDAAMKGMFIC 171
QY 204 YDMRPFELSLALAOAGAEILTPSARFSGITGPAHWEVL-----LRARAIET 249
DB 172 NDRMPEARVMGLRGAEILC-----GGYNTPTNPNPLVQHDHLTSFHHLLSMQAGSYON 226
QY 250 QCYVVAACGGRRHHEKRASTYGHSMVVDPMGTVAARCSE--GPGICLARIDILNYLRRL 308
DB 227 GAMSAAAGAGM--EENCMILGHSICVAPTEIYALTTLEDEVITAAVDDRCRELREHI 285
QY 309 PVF-OHRPDLG 320
DB 286 FNFQHRQPHYG 298

RESULT 6
US-08-876-398A-26
Sequence 26, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NAMBA, Hirokazu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FOLEY & LARDNER
;; STREET: 3000 K Street, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20007-5109

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/876,398A

;; FILING DATE: 16-JUN-1997

;; CLASSIFICATION: 435

;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/211,641

;; FILING DATE: 11-APR-1994

;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/JP93/01101

;; FILING DATE: 05-AUG-1993

;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: JP 340078/1992

;; FILING DATE: 21-DEC-1992

;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: JP 212692/1992

;; FILING DATE: 10-AUG-1992

;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wegner, Harold C.

;; REGISTRATION NUMBER: 25,258

;; REFERENCE/DOCKET NUMBER: 74129/130

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 672-5300

;; TELEFAX: (202) 672-5399

;; INFORMATION FOR SEQ ID NO: 26:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 303 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

US-08-876-398A-26

Query Match 11.4%; Score 199.5; DB 3; Length 303;
Best Local Similarity 26.9%; Pred. No. 9.3e-14;

Matches 83; Conservative 47; Mismatches 129; Indels 49; Gaps 14;

QY 48 LVAVCQ---VSTPRKQONFKTCALVREARLACLAFLPE--AFDFIAR---DPAET 98
DB 5 ILAVGGGPIARAEETREQVVRLLDMITKASRGANFIVPELATTFEPPRWHTDEAEL 64
QY 99 LHLSE-PLGGKLEETQALRECGMLSLGFGH---ERGQDWEQOKIYNCHVLLNSKA 154
DB 65 DSFEYTEMGPVVRPLPEKKAELGIGFNILGYAELVVEGG-----VKRFNFSILVDKSGK 119
QY 155 VVATYRRTKHLCDVEIPGGPWCESNSMPGSLSPVSTP-----AGKIGLAVC 203
DB 120 IYGYRKRIHL-----PGHK---EYEAYRPFQHLEKRYEPGDLGFPYVDVDAAMKGMFIC 171
QY 204 YDMRPFELSLALAOAGAEILTPSARFSGITGPAHWEVL-----LRARAIETOCYV 254
DB 172 NDRMPEARVMGLRGAEILCIGGYNTPTNPNPLVQHDHLTSFHHLLSMQAGSYONGAMSA 231
QY 255 AAACGGRRHHEKRASTYGHSMVVDPMGTVAARCSE--GPGICLARIDILNYLRRLPVF-Q 312
DB 232 AAGKVGH--EENCMILGHSICVAPTEIYALTTLEDEVITAAVDDRCRELREHI 290
QY 313 HRRPDLG 320
DB 291 HRPQPHYG 298

```

RESULT 7
US-08-876-3984-40
: Sequence 40 Application US/08876398A
: Patent No. 6083752
: GENERAL INFORMATION:
: APPLICANT: IKENAKA, Yasuhiro
: APPLICANT: NAKABA, Hirokazu
: APPLICANT: TAKANO, Masayuki
: APPLICANT: YAJIMA, Kazuyoshi
: APPLICANT: YAMADA, Yukio
: APPLICANT: TAKAMASHI, Satomi
: TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
: TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
: NUMBER OF SEQUENCES: 70
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FOLEY & LARDNER
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/876.398A
: FILING DATE: 16-JUN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/211,641
: FILING DATE: 11-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/JP93/01101
: FILING DATE: 05-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 340078/1992
: FILING DATE: 21-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 212692/1992
: FILING DATE: 10-AUG-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Wegner, Harold C.
: REGISTRATION NUMBER: 25,258
: REFERENCE/DOCKET NUMBER: 74129/130
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 672-5300
: TELEFAX: (202) 672-5399
: INFORMATION FOR SEQ ID NO: 40:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 303 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-876-3984-40

Query Match 11.4%; Score 199.5; DB 3; Length 303;
Best Local Similarity 26.2%; Pred. No. 9.3e-14;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

Dy 48 LVAVQ---VSTPRKQONFKTCALVVEAALGLACLAFLPE--AFDFLAR----DPART 98
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 ILAVQGGPIARAEFTREDOVAVRLDLMLTKAASRGANFTVPELALTFFPRNHTDEAL 64
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 99 LHLSE-PLSGKLLDEYUOLARECGMLSLGFGH---ERGODEQOTOKLYNCVLLNSKA 154
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 DSFYEETEMPGPVYRRLFEKAALGIGFNGLVAYELVVEG-----VKRRNTSLIVDKSG 119
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 155 VVATYRKTHLCDVLEIPGQCPMCSNSTMPGSLSPVSTP-----AGKIGLAVC 203
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 IVGKTKRTHL-----PGHK---EYEAVRPFQHLKKRYEPFGDLGPRVYDVDAKGMGTC 171
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 204 YDMPEPILSLAQAQGEILTYPSAFSGINCPAHWEVL-----LRAPIET 249
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 172 NDRMPEPEARMWRCMGRLGEIILC-----GCIYTPPHNPLVPPQDHILTSFHHLLSMQASQYN 226
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 250 QCYYVAAACGCRHNEKRAYSGHSMVNDPMTGVYARCSE-GRGLCLARIDNLYRQLRHL 308
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 GAWSAAGKAGKM-DENCMILGHSCIVAPTEGIYALTTLEDEVITAAVDDRCRELREHI 285
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 309 PVF-QHRRPDLTG 320
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 286 ENFKOHRQPOHYHG 298

RESULT 8
US-08-876-398A-44
; Sequence 44, Application US/08876398A
; Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NAKABA, Hirokazu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-876-398A-44

```

Best Local Similarity 26.28; Pred. No. 9.3e-14;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

[illegible]

RESULT 9
US-08-294-871A-6
Sequence 6, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiko
TITLE OF INVENTION: Immobilized Enzyme Preparation and
PROCESS FOR PRODUCING D-AMINO ACIDS
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER: PCT/JP91/01696
3      FILING DATE: 06-DEC-1991
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: JP 407922/1990
6      FILING DATE: 27-DEC-1990
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: JP 078840/1991
9      FILING DATE: 11-APR-1991
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: JP 140051/1991
12     FILING DATE: 12-JUN-1991
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER: PCT/JP92/00739
15     FILING DATE: 10-JUN-1992
16     PRIOR APPLICATION DATA:
17     APPLICATION NUMBER: JP 212692/1992
18     FILING DATE: 10-AUG-1992
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER: PCT/JP93/01101
21     FILING DATE: 05-AUG-1993
22     PRIOR APPLICATION DATA:
23     APPLICATION NUMBER: JP 340078/1992
24     FILING DATE: 21-DEC-1992
25     ATTORNEY/AGENT INFORMATION:
26     NAME: Wegner, Harold C.
27     REGISTRATION NUMBER: 25, 258
28     REFERENCE/DOCKET NUMBER: 74129/127/AOPA
29     TELECOMMUNICATION INFORMATION:
30     TELEPHONE: (202) 672-5300
31     TELEFAX: (202) 672-5399
32     TELEX: 904136
33     INFORMATION FOR SEQ ID NO: 6:
34     SEQUENCE CHARACTERISTICS:
35     LENGTH: 303 amino acids
36     TYPE: amino acid
37     TOPOLOGY: linear
38     MOLECULE TYPE: protein
39     US-08-294-871A-6

```

```

Query Match ..... 11.4% Score 198.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred No. 1.2e-13;
Matches 82; Conservativity 46; Mismatches 126; Indels 59; Gaps 14;

QY 48 IWAQCO---VTSPPKQONFKTCAELVREARLGAQLPE--AFDFIAR---DPAET 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 ILAVGQGGIARAETREQVAVRLDLMLTKAASRGNFIYFPELALTTFPPRHMFDEAL 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 LHLSE-PIGGKLLERYTQIARCGIAMLISGHH---ERQDMEQOKIYNCVILLNSKA 154
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 DSFETEMGGPVRRPLFEKAELIGFINGVAVELVEGG---VKRRPNTILVDKSK 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 VVAATYKTHLCDVEIIPGQGMPCSNSTMGPSLSPSVTP-----AGKITGLAVC 203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 120 IYGKTRKIH-----PGHK---EYAVRPFQHLERKYEPPGDLGPVVYDVDAKKMGFIC 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 YVMRPETSLALQAGAEITLTPSAFGSTITGAHKEVL-----LRARATET 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 172 NDRRPPEARVWGLGAELTIC-----GGYNTPTPNLVQNDHLTSFHHLSMKGSTYON 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 QCYVVAACCGRHNEKRASYSGHSMVVDPMGVAVARCSE-GPGICLARIIDLNTLRRL 308
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 227 GAMSAAAGVVG--EENCMLDGHSCIVAPGTGELVALTTTLEDEVITAAVYDLCRREI 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 PVF-QHRRDLYG 320
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 286 FNFQKHROPQHYG 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-08-294-871A-34
; Sequence 34, Application US/08294871A

```

Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/NOA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399

TELEX: 904136
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-34
Query Match 11.4%, Score 198.5, DB 2, Length 303,
Best Local Similarity 26.2%, Pred. No. 1.2e-13,
Matches 82, Conservative 46, Mismatches 126, Indels 59, Gaps 14,
OY 48 LVAQC---VSTPDKOQNFKTCAELVREARLGAFLAPE--AFDFIAR---DPAET 98
DB 5 ILAVGQGPRIARAEIREQVYVRLDMLTRFAASRGANFIYFPLALITFFPRRYFDEDEL 64
OY 99 LHLSE-PLGGKILEEYTLARECGIMLSIGFH---ERGQDMEOTOKIYNCHVLLNSKA 154
DB 65 DSFYETEMPGPVVRBLFEKAELGIGFNLGYALVEGQ---VKRRFNTSILVDSGK 119
OY 155 VVATYRKTHLCDVELIPGCGPMCESTNTPGSLSPVSTP-----AGKIGLAVC 203
DB 120 IYKTRKTHL---PGHK---EYDAVYRPOHLERKRYEPGDLGFPYVDVDAKMGMTIC 171
OY 204 YDMRPELSIALAQAGAEILTPSAFSGITGPAHNEVL-----LARALET 249
DB 172 NDRWRPEAMRVVGLGAEIIC-----GGYNTPTNPLVYQHDLHLSFHLISMQAGSYCN 226
OY 250 OCYVVAACGGRHNEKRASYGSHNVDPGTVVANCSE-GPGLCLARIDLNTLRRL 308
DB 227 GAMSAAAGVGM-EENCMLHSGCIYAPGELVATLTLEDEVITAAVDLDRCRELREH 285
OY 309 PVF-QHRRDLYG 320
DB 286 FNEKHROPQHYG 298
RESULT 11
US-08-294-871A-60
Sequence 60, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:


```

1 APPLICATION NUMBER: US 07/971,758
2 FILING DATE: 12-APR-1993
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 07/917,111
5 FILING DATE: 07-AUG-1992
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US 08/211,641
8 FILING DATE: 11-APR-1994
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: JP 400848/1990
11 FILING DATE: 07-DEC-1990
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: PCT/JP91/01696
14 FILING DATE: 06-DEC-1991
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: JP 407922/1990
17 FILING DATE: 27-DEC-1990
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: JP 078840/1991
20 FILING DATE: 11-APR-1991
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: JP 140051/1991
23 FILING DATE: 12-JUN-1991
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: PCT/JP92/00739
26 FILING DATE: 10-JUN-1992
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: JP 212692/1992
29 FILING DATE: 10-AUG-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: PCT/JP93/01101
32 FILING DATE: 05-AUG-1993
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: JP 340078/1992
35 FILING DATE: 21-DEC-1992
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Wegner, Harold C.
38 REGISTRATION NUMBER: 25,258
39 REFERENCE/DOCKET NUMBER: 74129/127/AOPA
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: (202) 672-5300
42 TELEFAX: (202) 672-5399
43 TELEX: 904136
44 INFORMATION FOR SEQ ID NO: 60:
45 SEQUENCE CHARACTERISTICS:
46 LENGTH: 303 amino acids
47 TYPE: amino acid
48 TOPOLOGY: linear
49 MOLECULE TYPE: protein
50
51 US-08-294-871A-60

```

```

Query Match 11.4% Score 198.5: DB 2: Length 303:
Best Local Similarity 26.2%: Pred. No. 1.2e-13:
Matches 82: Conservative 46: Mismatches 126: Indels 59: Gaps 14

OY 48 LVAVCQ---VSTPDKQONFKTCAELVREARLIGACLAFLPE--AFDTAR---DPAET 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 ILAVQCGPIRAAEETREQVVRLRLMLTGAARGANFIYFPELATLTFPRWYETDEAL 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 99 LHLSE-PLGSGKLEBYTOLARECGIWLSTGGFH---ERQDMDQPKIYNCHVLLNSKA 154
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 DSFYETMPGPPVVRRLFEEAALGICGNIGYALELYVEG-----YKRRENTSLVDKSGK 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 155 VVAIYRKTHLDCVEITPGGPMCESNSTMPGSLSESVSTP-----AGKIGLAVC 203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 IVGKRYRKITHL-----PGHK---EYAAVRFPHLEKRYEPFGDLGPVYDVDAKMMKMFIC 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 204 YDMRPPEISIALAOGAGELITPSPAFGSTITGAHMEVYL-----LRAAIEET 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 NDRMRPEAMRWGLNGALGAEITC-----GGYNTPTPHNLFVQHDHLLTSFHLLLSMQDSYON 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 250 OCYVVAACCGRHHEKRASYSGHSMVVDPMGTVAARCSE-GPGICLARIDANLYRLQRLRH 308

```

DB	QY	DB
227	309	286
GAMSAAPKSGM-EENCMLDGHSCIVAPTSYIALTTLEDEVTALTAVDRCRELRHNI	PVF-OHRRPDLYG	FNFKQHRQRPQHYG
285	320	298

```

1 RESULT 12
2 US-08-876-398A-6
3 : Sequence 6, Application US/08876398A
4 : Patent No. 6083752
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: IKENAKA, Yasuhiro
9 : APPLICANT: NAKABA, Hirokazu
10 : APPLICANT: TAKANO, Masayoshi
11 : APPLICANT: YAJIMA, Kazuyoshi
12 : APPLICANT: YAMADA, Yukio
13 : APPLICANT: TAKAHASHI, Satoshi
14 :
15 : TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
16 : THERMOSTABILITY AND USE THEREOF
17 :
18 : NUMBER OF SEQUENCES: 70
19 :
20 : CORRESPONDENCE ADDRESS:
21 : ADDRESSEE: FOLEY & LARDNER
22 : STREET: 3000 K Street, N.W.
23 : CITY: Washington
24 : STATE: D.C.
25 :
26 : COUNTRY: U.S.A.
27 : ZIP: 20007-5109
28 :
29 : COMPUTER READABLE FORM:
30 : MEDIUM TYPE: Floppy disk
31 : COMPUTER: IBM PC compatible
32 : OPERATING SYSTEM: PC-DOS/MS-DOS
33 : SOFTWARE: PatentIn Release #1.0, Version #1.30
34 :
35 : CURRENT APPLICATION DATA:
36 : APPLICATION NUMBER: US/08/876.398A
37 : FILING DATE: 16-JUN-1997
38 : CLASSIFICATION: 435
39 :
40 : PRIOR APPLICATION DATA:
41 : APPLICATION NUMBER: US 08/211,641
42 : FILING DATE: 11-APR-1994
43 :
44 : PRIOR APPLICATION DATA:
45 : APPLICATION NUMBER: WO PCT/JP93/01101
46 : FILING DATE: 05-AUG-1993
47 :
48 : PRIOR APPLICATION DATA:
49 : APPLICATION NUMBER: JP 340078/1992
50 : FILING DATE: 21-DEC-1992
51 :
52 : PRIOR APPLICATION DATA:
53 : APPLICATION NUMBER: JP 212692/1992
54 : FILING DATE: 10-AUG-1992
55 :
56 : ATTORNEY/AGENT INFORMATION:
57 : NAME: Wegner, Harold C.
58 : REGISTRATION NUMBER: 25,258
59 :
60 : REFERENCE/DOCKET NUMBER: 74123/130
61 :
62 : TELECOMMUNICATION INFORMATION:
63 : TELEPHONE: (202) 672-5300
64 : TELEFAX: (202) 672-5399
65 :
66 : INFORMATION FOR SEQ ID NO: 6:
67 : SEQUENCE CHARACTERISTICS:
68 : LENGTH: 303 amino acids
69 : TYPE: amino acid
70 : TOPOLOGY: linear
71 :
72 : MOLECULE TYPE: protein
73 :
74 : US-08-876-398A-6

```

```

Query Match Similarity      11.4%  Score 198.5  DB 3:  Length 303:
Best Local Similarity      26.2%  Pred No. 1.2e-13:
Matches      82:  Conservative  46:  Mismatches 126:  Indels  59:  Gaps  14:

48 LVAVCQ--VTSPTKQCNFKTCAGELVREARIGACLAFLPE--AFDFTAR---DPAET  98
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  5 ILAVCGGFIARAEETREGVVRLMLMTKTAASRGAGNFIETPLATLTERRRHHFDEAL  64
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

QY	99	LHLSP-PLGGKLTLEBYOPLAEECJMLTSLGGFH---	ERGDWQETOKIYNCBYLNSKGA	154
Db	65	DSFYETEMPGVYVPRLEPKAAELCIGNLGTAELVYBGG---	VKRENTSTLIVDWSGK	119
QY	155	VVATYRKTHLCLDVEILPGGSPMCESNSTMPGSLSPVSTP-----	AGKIGLAVC	203
Db	120	IVGKRIKTHL-----PGHK---EYEAVRPFQHLERKFFEEPGDLGFPYVDVDAKMGMFIC	171	
QY	204	YDMRPPELSTLAAQGAELIYTPAFSGSTNGPAHWEL-----	LRARLET	249
Db	172	NDRKRPPELWRVWGLRGAFILC-----GGYNTPTNPILPQCHDLTSHHLLSMQAGSYQN	226	
QY	250	QCYYVAAAQCSRHNEHKRASYSCHSMVDPWGTVAARCSE-GRGICLARIDILNYROLRIHL	308	
Db	227	GAMSAAPAAKVGM-EENCMLTGHSCIVAPRTGIVALTTLTLEDEVITAAVDLDRCRELREHI	285	
QY	309	PVF--OHRRPDLXG	320	
Db	286	FNEKOHROPQIHG	298	

RESULT 13
 US-08-876-398A-34
 : Sequence 34, Application US/08876398A
 : Patent No. 6083752
 : GENERAL INFORMATION:
 : APPLICANT: IKENAKA, Yasuhiro
 : APPLICANT: NAKABA, Hirokazu
 : APPLICANT: TAKANO, Masayuki
 : APPLICANT: YAJIMA, Kazuyoshi
 : APPLICANT: YAMADA, Yukio
 : APPLICANT: TAKAHASHI, Satomi
 : TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
 : TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
 : NUMBER OF SEQUENCES: 70
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: FOLEY & LARDNER
 : STREET: 3000 K Street, N.W.
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: U.S.A.
 : ZIP: 20007-5109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/876,398A
 : FILING DATE: 16-JUN-1997
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/211,641
 : FILING DATE: 11-APR-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: WO PCT/JP93/01101
 : FILING DATE: 05-AUG-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 340078/1992
 : FILING DATE: 21-DEC-1992
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 212692/1992
 : FILING DATE: 10-AUG-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Wegner, Harold C.
 : REGISTRATION NUMBER: 25,258
 : REFERENCE/DOCKET NUMBER: 74129/130
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 672-5300
 : TELEFAX: (202) 672-5399
 : INFORMATION FOR SEQ ID NO: 34:
 : SEQUENCE CHARACTERISTICS:

```

; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-876-398A-34

```

Query Match	11.48;	Score 198.5;	DB 3;	Length 303;
Best Local Similarity	26.28;	Pred. No. 1.2e-13;		
Matches	82;	Conservative	46;	Mismatches 126;
			Indels	59;
			Gaps	14;

Oy	48	LWAOO---	YVSTDKOONFTKCAEILVEAKRIADSCAFIDE--	APFQIR----	DAET	98
Dd	5	ILAVGOORIRARETREVUVUVRLLDMLTKAASRANRTIPELALTTFFPRMYFTDBAEI.				64
Oy	99	LHLSE-PLGSKILLEBYTOIANECGLIMJSLGCFH--	ERGDMEOTOKIYCNCYLINSKA			154
Dd	65	DSEFYETEMRGVPVRLPDLFEKAAELDIGNLSGAELVVBGG-----	VKRFRNTSILYVDSGK			119
Oy	155	VVAATRKYNILDVEITPCGGPMSCSNSTPGFSLESPVSTP-----	ACKIGLAVC			203
Dd	120	IUGVKRIHTL-----	PGHK-----	EYEAVRPQGNLEKYFERFDLGFPVUYDVDAKMGWFIC		171
Oy	204	YDMRPBELSLADAOGAEILTTPRAFSGITGRAMEVL-----	LRAKIET			249
Dd	172	NDRMRPEWMRYMGLRGAIIC-----	GGYNTPTNPVLPONDILTSFNHLLSMOAGSYQN			226
Oy	250	OCYUVAALAOCSRNHEKRAYSHGVNDPVMTCTVAVRCSE-GRGLCLARIDILNYROLREHNL				308
Dd	227	GAMSAAGAUKVM-EENCHMLGHOSIVARTGBVALTTLTLEDEVITTAUVDDRCRELYEHNI				285
Oy	309	PVF-QHRRPDLXG	320			
Dd	286	FNEKHROPDHNG	298			

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

? APPLICATION NUMBER: JP 340078/1992
? FILING DATE: 21-DEC-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 212692/1992
? FILING DATE: 10-AUG-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Wegner, Harold C.
? REGISTRATION NUMBER: 25,258
? REFERENCE/DOCKET NUMBER: 74129/130
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 672-5300
? TELEFAX: (202) 672-5399
? INFORMATION FOR SEQ ID NO: 60:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 303 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-876-398A-60

Query Match      11.4%; Score 198.5; DB 3; Length 303;
Best Local Similarity 26.2%; Pred. No. 1.2e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 48 LVAVCO---VTSPPDKOQNKTKCAELVREARLGACLAFLPE--AFDFIAR----DPAET 98
DB 5 ILAVGGOGPIARAETREQVVVRLDMLTKAASGANFIYFPELALTTFFPRMHTFDEAEL 64
QY 99 LHLSE-PLGSKLLEYTQLAKECGMLSLGFFH---ERGDWEQOTOKIYNCHVLNLSKGA 154
DB 65 DSEYETEMPGVVRPLFEKKAELGIGFNIGYALVVEG-----VKRRFTSLIVDKSGK 119
QY 155 VVATYKTHLCDEVEITGQGPMSNSTMGPSLESFVSTP-----ACKIGLAVC 203
DB 120 IYKGYRKIHL-----PGHK-----EYEAIVRPFQHLKRYFEFGDLGFPYVDVDAKMGWFTIC 171
QY 204 YDMRFPELSIALQAAGAEILTYPSAFSGITGPAHWEVL-----LRARALET 249
DB 172 NDRRWPEARVWKLKGAELIIC-----GYNTPYHNPLVPOHDLTSEHILLSMQAGSYON 226
QY 250 QCVVVAACGGRHHERKASYSVMVDPMGTVVARGSE-GPGICLARIDNLYRQLRHL 308
DB 227 GAMSAAAGKSGM-EENCMLGHSCIVAPGEIYALTTLEDEVITAAVDLDRRELREHI 285
QY 309 PVF-QHRPDLVG 320
DB 286 FNFKHROPQHYG 298

RESULT 15
US-08-294-871A-62
? Sequence 62, Application US/08294871A
? Patent No. 5824522
? GENERAL INFORMATION:
? APPLICANT: Ikenaka, Yasuhiro
? APPLICANT: Nanba, Hirokazu
? APPLICANT: Takano, Masayuki
? APPLICANT: Yajima, Kazuyoshi
? APPLICANT: Yamada, Yukio
? APPLICANT: Takahashi, Satomi
? APPLICANT: Okubo, Kazuma
? APPLICANT: Yamada, Kazuhiko
? APPLICANT: Hirataishi, Yoshio
? TITLE OF INVENTION: Immobilized Enzyme Preparation and
? TITLE OF INVENTION: Process for Producing D-a-Amino Acid
? NUMBER OF SEQUENCES: 70
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Foley & Lardner
? STREET: 3000 K Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.
```

```

? ZIP: 20007-5109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/294,871A
? FILING DATE: 22-AUG-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/971,758
? FILING DATE: 12-APR-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/917,111
? FILING DATE: 07-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/211,641
? FILING DATE: 11-APR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 400848/1990
? FILING DATE: 07-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP91/01696
? FILING DATE: 06-DEC-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 407922/1990
? FILING DATE: 27-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 078840/1991
? FILING DATE: 11-APR-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 140051/1991
? FILING DATE: 12-JUN-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP92/00739
? FILING DATE: 10-JUN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 212692/1992
? FILING DATE: 10-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP93/01101
? FILING DATE: 05-AUG-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 340078/1992
? FILING DATE: 21-DEC-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Wegner, Harold C.
? REGISTRATION NUMBER: 25,258
? REFERENCE/DOCKET NUMBER: 74129/127/AOPA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 672-5300
? TELEFAX: (202) 672-5399
? TELEX: 904136
? INFORMATION FOR SEQ ID NO: 62:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 303 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-294-871A-62

Query Match      11.3%; Score 197.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred. No. 1.5e-13;
Matches 82; Conservative 45; Mismatches 127; Indels 59; Gaps 14;

QY 48 LVAVCO---VTSPPDKOQNKTKCAELVREARLGACLAFLPE--AFDFIAR----DPAET 98
DB 5 ILAVGGOGPIARAETREQVVVRLDMLTKAASGANFIYFPELALTTFFPRMHTFDEAEL 64
QY 99 LHLSE-PLGSKLLEYTQLAKECGMLSLGFFH---ERGDWEQOTOKIYNCHVLNLSKGA 154
```

```

Db 65 DSFYETEMPGPVVRPLFEKAEIIGIFNLGYAELVVEG-----VKRRPNTSILYDKSGK 119
Oy 155 VVATYRKTHLCOVEIIPGOGPMCESNSTMPGSPLESVPSTP-----AGKTGLAVC 203
Db 120 IVGKTRKTHL-----PGHK---EYEAIRPFQHLKRYEFEPDGLGFPVYDVDAKMGMTIC 171
Oy 204 YDMRPEPELSLALQAAGAEILTYPSAFGSITGPAHWEVL-----LRARALET 249
Db 172 NDRMRPEARVAVMGLRGAELIC-----GGYNTPTHNPAPVQHDHLTSFHHLLSMQAGSYGN 226
Oy 250 OCYVVAACCGRHHEKRASYGSHMVDPMGTVVARCSE-GPGICLARIDNLTNRQLRHL 308
Db 227 GAWSAAGVGM-EENCMLDGHSCIYAPTGELVALTTLEDEVITAAYVDLDRCRELREHI 285
Oy 309 PVF-OHRRPDLTG 320
Db 286 FNFKQHRQPOHYG 298

```

Search completed: April 26, 2002, 16:55:11
 Job time: 254 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:49 ; Search time 21.41 Seconds

(without alignments)
553.141 Million cell updates/sec

Title: US-09-357-675C-22

Perfect score: 1736

Sequence: 1 MGFTRPPHQLCTGYRL.....LPVQRHRRPDLVGSIGHPLS 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	523	30.1	307	1 YJM6_YEAST	P47016 saccharomyc
2	492	28.3	272	1 Y601_SYNY3	P55175 synchocyst
3	437.5	25.2	322	1 YAUB_SCHPO	Q10166 schizosacch
4	419	24.1	291	1 YL85_YEAST	P49954 saccharomyc
5	364	22.0	262	1 YBEM_ECOLI	P39874 escherichia
6	361	20.8	262	1 YBEM_ECO57	P58054 escherichia
7	268.5	15.5	340	1 Y480_MYCTU	Q11146 mycobacteri
8	268	15.4	285	1 YPOQ_PSEFL	P55176 pseudomonas
9	249.5	14.4	261	1 YAG5_STRAU	P55177 staphylococ
10	237.5	13.7	234	1 YAG5_STRAU	P55178 staphylococ
11	206	11.9	346	1 NR1L_ARATH	P32861 arabidopsi
12	203	11.7	346	1 NR1L_ARATH	P32861 arabidopsi
13	193	11.1	393	1 BOP_RAT	P46010 arabidopsi
14	187.5	10.8	339	1 NR1L_ARATH	Q03248 ratius norv
15	186.5	10.7	346	1 ALAM_PSEAE	P11436 pseudomonas
16	181.5	10.5	513	1 YHXC_BACSU	P54608 bacillus su
17	170	9.8	355	1 NR1L_ARATH	P46011 arabidopsi
18	168.5	9.7	345	1 ALAM_RHOER	Q01360 rhodococcu
19	156	9.0	349	1 NR1L_TOBAC	Q42965 nicotiana t
20	148	8.5	256	1 YAEV_ECOLI	Q47679 escherichia
21	143	8.2	21	1 NAE2_THEMA	Q9X0Y0 thermotoga
22	126	7.3	356	1 NR1L_ALCPA	P20960 alcaligenes
23	123.5	7.1	567	1 MADE_AOUAE	O67091 aquifex aeo
24	117.5	6.8	368	1 CYH4_GLOSO	P32964 glaucocercos
25	112.5	6.5	199	1 Y1Q4_YEAST	P40447 saccharomyc
26	111.5	6.4	552	1 MADE_RHOCA	Q03638 rhodobacter
27	108	6.2	519	1 LNT_SYNY3	P74055 synchocyst
28	107	6.2	383	1 NR1L_RHOER	Q02068 rhodococcu
29	106.5	6.1	679	1 MADE_MYCTU	P71911 mycobacteri
30	100	5.8	794	1 SEIL_HUMAN	Q929V2 homo sapien
31	99	5.7	541	1 LNT_CHLPL	Q92791 chlamydia p
32	98.5	5.7	511	1 LNT_PSEAE	Q92186 pseudomonas
33	97.5	5.6	365	1 NR1L_RHOER	Q03217 rhodococcu

34	96.5	5.6	680	1 MADE_MYCLE	Q9cbz6 mycobacteri
35	95	5.5	512	1 VNN1_MOUSE	Q920k8 mus musculu
36	93.5	5.4	455	1 TMS5_MOUSE	Q9er04 mus musculu
37	93	5.4	425	1 YGEP_ECOLI	P52048 escherichia
38	92.5	5.3	542	1 LNT_CHLTR	O84539 chlamydia t
39	92	5.3	790	1 SEIL_MOUSE	Q92296 mus musculu
40	91.5	5.3	513	1 VNN1_HUMAN	O95497 homo sapien
41	91.5	5.3	520	1 VNN2_HUMAN	O95498 homo sapien
42	90.5	5.2	541	1 LNT_RHIME	O52910 rhizobium m
43	90	5.2	349	1 NR1L_KLEPO	P10045 klebsiella
44	87.5	5.0	496	1 LNT_RICPR	Q92d93 rickettsia
45	86.5	5.0	500	1 VNN3_MOUSE	Q9qz25 mus musculu

ALIGNMENTS

RESULT	ID	YJM6_YEAST	STANDARD	PRT	307 AA.
AC	P47016				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	HYPOTHEICAL 34.7 KDA PROTEIN IN SPT10-GCD14 INTERGENIC REGION.				
GN	YJL126W OR J0706.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=5288C / FY1679;				
RA	MEDLINE=97103775; PubMed=8948101;				
RX	Czilepuch C., Kordes E., Pujol A., Jauniaux J.-C.;				
RT	"Sequencing analysis of a 40.2 kb fragment of yeast chromosome X				
RT	reveals 19 open reading frames including URA2 (5' end), TRX1, PBS2,				
RT	SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCR7, GZF3, two tRNA genes,				
RT	three remnant delta elements and a TY4 transposon."				
RL	Yeast 12:1471-1474(1996).				
CC	-1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: Z49401; CAA89421.1; -				
DR	SGD: S0003662; YJL126W.				
DR	InterPro: IPR0031010; CN_hydrolase.				
DR	InterPro: IPR001110; UPF0012.				
DR	Pfam: PF00795; CN_hydrolase: 1.				
DR	PROSITE: PS01227; UPF0012: 1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 307 AA: 34693 MW: 48787CA3B10A828E CRC64:				
Query Match	30.1%: Score 523; DB 1; Length 307;				
Best Local Similarity	38.3%: Pred. No. 5e+40;				
Matches 116; Conservative 55; Mismatches 104; Indels 28; Gaps 6;					
OY	41 ELPLVAVCVTSPFNKQENFKCAELVQEARLACIAFLPEAFDFTARNPAETILL-- 97				
DB	4 KLRKVAQAQGLSSADLTKNLKLVKVELISEAQKADAVFLPEASDYISQNLHSRYLAOK 63				
OY	98 SEPLNGLLCOYSLARE--CGIWLSLG-GFHEGQD-WEDONQKTYNCVILNKSQSYVA 153				
DB	64 SPKFIROILOSITDVLVDNSRNIDVSGIVHLPSQDILLENDVRVNVLLYIDHEGRTIQ 123				
OY	154 SYRTHLCDEVEIPQGMRESNRYTKPGGTLEPPYKTAGKGLAICVDMRPEELSLKLAQ 213				

ID	SYN3	STANDARD	PRT	272 AA
Db	124	EXOKHLHLDVDPVNPGRPIILAKESKSYVPGCAIPDIIESPLKLSALCYDIDRPFESLKLRS	183	
Qy	214	AGAEILTYPSAFSGVTPGAHWEVILLRARIIESQCYVIAAOCGRH-----	258	
Db	184	MGAEILCPSPAFATIKTGBAHEMLLGRARAVDPQCYYVLMGQYGMHDLSDPEWEKQSHMSA	243	
Qy	259	---HETRASYSGSMVVDWGVYVAVC---SEPGCLCARIDHAFIQOMROHLPVQHRRP	312	
Db	244	LEKSRRSRWGSHMVIDPWGKIHAADPSYVGPOLILADLRELLQETRNKPMWNORD	303	
Qy	313	DLV 315		
Db	304	DLF 306		
RESULT	2			
ID	Y601_SYNN3			
AC	P55175	STANDARD	PRT	272 AA
DT	01-OCT-1996	(Rel. 34, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	HYPOTHETICAL 30.2 KDA	PROTEIN SLI0601		
GN	SLI0601			
OS	Synechocystis sp. (strain PCC 6803)			
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis			
OX	NCBI_TaxID=1148;			
RA	SEQUENCE FROM N.A.			
RP	MEDLINE=96127529; PubMed=8590279;			
RA	Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,			
RA	Sugita M., Tabata S.;			
RT	"Sequence analysis of the genome of the unicellular cyanobacterium			
RT	Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb			
RL	region from map positions 64% to 92% of the genome."			
RL	DNA Res. 2:153-166(1995).			
CC	-I- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: D64002; BAA10370.1; -			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	InterPro: IPR001110; UPF0012.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
KW	Hypothetical protein; Complete proteome.			
SO	SEQUENCE 272 AA; 30191 MW; 7787BA9BZBE003A1 CnC64;			
Query Match	28.3%; Score 492; DB 1; Length 272;			
Best Local Similarity	41.6%; Pred. No. 2.7e-37;			
Matches 114; Conservative 38; Mismatches 106; Indels 16; Gaps 5;				
Qy	44	LVAACQVSTPNKQENFKTCALVEOEAARLGACIAFLPAEDFARNPAETLLISEPLNG	103	
Db	5	LAALAGMTSRPLTENILDEAELLIDLAROGAEIYGLPENFAFLG-NETEKLEQATATAT	63	
Qy	104	DLLGQISOLARCGTWSLGRF-----HERGQDWQONQIKYCNHVLNLSKGSVVAITYKT	158	
Db	64	ATEKFLQTMADQFYVTIILAGGPFPPVAGEAG-----KAYNTATLIPNGELARIRHY	116	
Qy	159	HLCDVEIFGQGGRRSNYTKPGCTLEPV--KTPAGKYGLATCYMRPEPLSLKLAQGA	216	
Db	117	HLFDVNVDPGNTYMS-ATVAAQGXKPPVYHSDSGNIGLSICYVRPEPLRYLSTRQGA	175	
Qy	217	ELIYPSAFSGVTPGAHWEVILLRARIIESQCYVIAAOCGRHNETRASYSGSMVVDWVPGCT	276	

DB	176	DVLEPRAAFATYATGODHNOVLQALAIENITCYVIAIPATGCHYERHRTNHGAMITDPMGV	235
QY	277	VVARCEGPGCLARIDLHFLQOMRQHLPVFOHR	310
Db	236	ILADAGEKPLAIAEINPDRLKOVNQMPISQHR	269
RESULT	3		
YAUB_SCHPO	STANDARD;	PRT;	322 AA.
ID	YAUB_SCHPO		
AC	Q10166;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	HYPOTHEETICAL 35.7 KDA PROTEIN C26A3.11 IN CHROMOSOME I.		
GN	SPAC26A3.11.		
OS	Schizosaccharomyces pombe (fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
CC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
CC	Schizosaccharomyces.		
OX	NCBI_TaxID=4896;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-972;		
RA	McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;		
RL	Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; 269240; CAAG3234.1; -		
DR	InterPro: IPR003010; CN_hydrolase.		
DR	InterPro: IPR001110; UPF0012.		
DR	Pfam: PF00795; CN_hydrolase; 1.		
DR	PROSITE: PS01227; UPF0012; 1.		
KW	Hypothetical protein.		
SO	SEQUENCE 322 AA; 35679 MW; EA6F39B160C7F49F CRC64;		
Query Match	25.2%; Score 437.5; DB 1; Length 322;		
Best Local Similarity	33.4%; Pred. No. 2.9e-32;		
Matches 102; Conservative	51; Mismatches 113; Indels 39; Gaps 5;		
QY	34	MSSSTSWELP-----LVAVCQVSTPNKQENFKTCAELVQEAARLGAFLPEAFDFI	87
Db	29	MSVSASSLVPRDFRAFRIGLVOLANTKQKSENLQARLKVLEAAKNSNVIVLPEIF--	85
QY	88	ARNPAETLLLSLEPLNGDLGQY-----SOLARECGIWLSLSGFHERGOWE	133
Db	86	-----NSPYGTGYFNQYAEPIEESSPSYQALSSMAKDTKYLLTGGSIPER-----	130
QY	134	ONQKTYNCHVILNMSGSVASVYARKTHLCDVLEPPGGGPMRESNYYTRPGGTLEPVTYPAGK	193
Db	131	KDGRKYNTAMAFVDPBGKLIIVARRKIHLPDIDIPGGVSVFRESDSUSPGAM--TMDTEYK	189
QY	194	VGLAICYMPRPPELSLKLAQGAELITFPASFGSVTGAHNEVLLRAAIESOCVIAA	253
Db	190	FGLGICYDIRPEPLAMIAARNGCSMTIYPGAFLNSTGPLHVELLARARAVDNEMFVACCA	249
QY	254	QCGRHETRASVYGHSMVVDPMGVVAVARCESEPGCLARIDLHFLQOMRQHLPVFOHRPD	313
Db	250	PARDNMADYHSMGHSTVVDPGKVATATDEKPSIYVADIDPSVNSTANSVPITYQRFED	309
QY	314	LYGSL 318	
Db	310	YSEV 314	

```

RESULT 4
YL85_YEAST STANDARD; PRT; 291 AA.
ID YL85_YEAST
AC P49554;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOHETICAL 32.5 KDA PROTEIN YLR351C.
GN YLR351C OR L9638.5.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hallier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Mian M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U19102; AAB67751.1; -
CC SGD: S0004343; YLR351C.
CC InterPro: IPR003010; CN_hydrolyase.
CC InterPro: IPR001110; UPF0012.
CC Pfam: PF00795; CN_hydrolyase; 1.
CC PROSITE: PS01227; UPF0012; 1.
CC KW Hypothetical protein.
SQ
SEQUENCE 291 AA; 32549 MW; AB13744120088827 CRC64;

```

```

Query Match 24.1%; Score 419; DB 1; Length 291;
Best Local Similarity 33.8%; Pred. No. 1.2e-30;
Matches 99; Conservative 48; Mismatches 110; Indels 36; Gaps 6;

```

```

QY 45 VAVCVQVT-STPKQKNFTCAELVQEAR--LGACIAFPPEAFDPLARPAETLLISEPL 101
DB 13 VALVOLSSSSDPKMANLQRAATFIERANKEDPTRLVLPPECF-----NSPY 59
QY 102 NGDLLGOYSQ-----LAECGIMTSLGFFHFGQDMEQNOKINYCNVLT 145
DB 60 STDQPKRSEVINKEPSTSVQVFLSNLANKKEIILVGGTIPLED---PRTKTIYNTSIIF 116
QY 146 NKSQSVASVYRKTLCADVEIPQGGPMRESNYPKGTLEPPVKTAGKGLAICYDMRPP 205
DB 117 NEDGKLIDKHRRVHLDEVDIPMGISFHESETLSP-GEKSTITDTKYKGKGVGICYDMRPP 175
QY 206 ELSLKLAQAGAEILTPPAFGSVTPPAHWEVLLARAITSQCYVIAAOCGRHETIRASY 265
DB 176 ELAMLSAKKGAFAAMITYPAPFNTVITGRLHMLLARSRAVDNOYVIMCSPARNLQSSYHAY 235
QY 266 GHSWVDVPMGCVVAVRCSEPGCLARIDLHFLQQRKHLPVFOHRRPDDYGSU 318
DB 236 GHSIVYDPRKRTIVALEAGCEELITVLELDPEVLESFRQAVPLTKQRFDYISDV 288

```

```

RESULT 5
YBEM_ECOLI

```

```

ID YBEM_ECOLI STANDARD; PRT; 262 AA.
AC P39874; P77192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL PROTEIN YBEM.
GN YBEM OR B0625/B0626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alpa H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sameshi G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RL "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-51 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95075315; PubMed=7984109;
RA Yamanaka K., Milani T., Ogura T., Niki H., Hiraga S.;
RT "Cloning, sequencing, and characterization of multicopy suppressors
RT of a mukB mutation in Escherichia coli.";
RL Mol. Microbiol. 13:301-312(1994).
RN [5]
RP SEQUENCE OF 31-78 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Robison K., O'Keefe T., Church G.M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [6]
RP CONCEPTUAL TRANSLATION.
RA Budd K.E.;
RL Unpublished observations (JAN-2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -1- SIMILARITY: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC POSITION 67.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000167; AAC73726.1; ALT_FRAME.
CC DR EMBL: AE000167; AAC73727.1; ALT_FRAME.
CC EMBL: D90703; BAA35268.1; ALT_FRAME.

```


RA Novick R.P., Projan S.J., Kornblum J., Ross H.F., Ji G.,
 RA Kreiswirth B., Vandenesch F., Moghazeh S.;
 RT "The agr P2 operon: an autocatalytic sensory transduction system in
 RT *Staphylococcus aureus*.";
 RL Mol. Gen. Genet. 248:446-458(1995).
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X52543; CAA3679.1; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 KW Hypothetical protein.
 SO SEQUENCE 261 AA; 29826 MW; 2303D95A8C18F838 CRC64;

Query Match 14.4%; Score 249.5; DB 1; Length 261;
 Best Local Similarity 27.1%; Pred. No. 2.3e-15;
 Matches 79; Conservative 51; Mismatches 109; Indels 53; Gaps 9;
 QY 40 WEPLV-----AVCYSTRPNKQENKTCALVQEARLGCALFPE-----AFD 85
 Db 6 YQLPYVGGSSKNETQTTFQWFEKNMAEVDVYVLPENMNNYDLEHINEKADNNGSFS 65
 QY 86 FIANPFAETLLSEPLNGDLL-GQYSQLARECGIWLISGFERHGDQWONKTYCHVL 144
 Db 66 FIKH-----LAKKYVDIYAGSVSN-----RNNQIFNTAFS 97
 QY 145 LNSKGSVAVSYRKTHTLQDVEIPGQPMRESNYTRPGGLEPVPKPAKG-VGLAICYDMR 203
 Db 98 VNSGQLINEYDKVHLVPM-----LREHEFLTAGEVVAEPFLSDQTYTQLICYDLR 150
 QY 204 FPPELSKLAAQAGAILTPPSAFSGSVTPGAHWEVLLRAATISQCYVIAAOCGRHETRA 263
 Db 151 FPELRLYPASSGAKIAYVQWQ-PMRSLQHWHSLLKARATENNMFVIGTNGFGDGTNEY 209
 QY 264 SYGSHVVDPMGVAVARCSEPGCLARIDLFQOMKROHLPVPOHRRPDL 315
 Db 210 A-GHSIYINRGDLVGLNESADILTYVDLINEVEQORENIPVKSIXLDL 260

RESULT 10
 YAG5_STALU STANDARD; PRT; 234 AA.
 AC P55178;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOHETICAL PROTEIN IN AGR OPERON (ORF 5) (FRAGMENT).
 OS *Staphylococcus lugdunensis*.
 OC Bacteria; Firmicutes; Bacilli; Clostridium group;
 OC *Bacillus/Staphylococcus* group; *Staphylococcus*.
 OX NCBI_TaxID=28035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9336158; Pubmed=8359673;
 RA Vandenesch F., Projan S.J., Kreiswirth B., Etienne J., Novick R.P.;
 RT "Agr-related sequences in *Staphylococcus lugdunensis*.";
 RL FEMS Microbiol. Lett. 111:115-122(1993).
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L13334; AAA71975.1; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SO SEQUENCE 234 AA; 26495 MW; C1CBADCA5E1389A7 CRC64;

Query Match 13.7%; Score 237.5; DB 1; Length 234;
 Best Local Similarity 28.7%; Pred. No. 2.5e-14;
 Matches 74; Conservative 37; Mismatches 118; Indels 29; Gaps 6;
 QY 59 NFKTCALVQEARLGCALFPEAFDIARNPAETLLSEPLNGDLIGYSQLARECGI 118
 Db 4 NKQTDVVLTPMMNNGVALLEQLEKADFLERSDFI-----KNLALQYQV 49
 QY 119 WLSIGFERHGDQWONKTYCHVLLNSKGSVAVSYRKTHTLQDVEIPGQPMRESNYTK 178
 Db 50 DIAGSVSNKHH-----HIFNTAFIDKTKVINOYDKMHLVPM-----LDEPAFLT 97
 QY 179 PGCTLEPVPKPAKG-KVGLAICYDMRPELSKLAAQAGAILTPPSAFSGVTGAHWEV 237
 Db 98 AGKVPPEFKLSNGVKTQYQMLRPFELRLYPARSATITAFVVAQMPASR-LNHQVYL 156
 QY 238 LRAARISQCYVIAAOCGRHETRAVSGHVVDPWGTAVARCSEPGCLARIDLF 297
 Db 157 LKARATENNMFVIGTNGFGDGTNEYA-GHSVAIINPGEIITQELSTKEKELTVIIDDV 215
 QY 298 QOMROHLPVPOHRRPDL 315
 Db 216 EQORKAIPVFDLSLVPHLY 233

RESULT 11
 NRLL_ARATH STANDARD; PRT; 346 AA.
 AC P32961; Q42543; Q04908;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NITRILASE 1 (EC 3.5.5.1).
 GN NIT1 OR AT3G44310 OR T10D17_100.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RP STRAIN=CV. LANDSBERG ERECTA; TISSUE=leaf.
 RX MEDLINE=92209532; Pubmed=155601;
 RA Bartling D., Seedorf M., Mithofer A., Weller E.W.;
 RT "Cloning and expression of an Arabidopsis nitrilase which can convert
 RT indole-3-acetonitrile to the plant hormone, indole-3-acetic acid.";
 RL Eur. J. Biochem. 205:417-424(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Zhou L., Bartel B., Thorburn R.W.;
 RT "Nucleotide sequence of the Arabidopsis thaliana nitrilase 1 gene.";
 RL (in) Plant Gene Register PGR95-130.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98145459; Pubmed=9484465;
 RA Hillebrand H., Bartling D., Weller E.W.;
 RT "Structural analysis of the nit2/nit1/nit3 gene cluster encoding
 RT nitrilases, enzymes catalyzing the terminal activation step in indole-

```

RT acetic acid biosynthesis in Arabidopsis thaliana."
RL Plant Mol. Biol. 36:89-99(1998).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unsel'd M.,
RA Fattmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Deleney M., Boutry M., Griwell L.A., Maché R., Pulgomech P.,
RA De Simone V., Choise N., Artiguenave F., Robert C., Brotier P.,
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Mumbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T., Nordstiek B.,
RA Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,
RA Cooke R., Laurie M., Berger-Llauro C., Purnelle B., Masny D.,
RA de Haan M., Maare A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argitlon A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Uteback T., Fujii C.Y., Shea T.P.,
RA Cressy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltscher J., Sellers P., Gill J.E., Feldlyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Friess C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsuno M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
RA Maranabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
RN [5]
RN CHARACTERIZATION
RN STRAIN=CV, LANDSBERG, ERCTA;
RX MEDLINE=94286570; PubMed=8016109;
RA Bartling D., Seedorf M., Schmidt R.C., Weiler E.W.;
RT "Molecular characterization of two cloned nitrilases from Arabidopsis
RT thaliana: key enzymes in biosynthesis of the plant hormone indole-3-
RT acetic acid."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
RN [6]
RN FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
RN INDOLE-3-ACETIC ACID.
RN CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
RN DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, BUT AT A
RN VERY LOW LEVEL DURING THE FRUITING STAGE.
RN SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN -----
DR EMBL: X63445; CAA45041.1; -
DR EMBL: U38845; AAB05221.1; -
DR EMBL: Y07648; CAA68935.2; -
DR EMBL: AL353865; CAB88999.1; -
DR PIR: S22398; S22398.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR000132; Nitril_cyn_hydrolase.
DR Pfam: PF00795; CN_hydrolase_1.
DR PROSITE: PS00920; NITRIL_CHT_1; 1.
DR PROSITE: PS00921; NITRIL_CHT_2; 1.
DR Hydrolase: Multigene family.
KW ACT_SITE 186 186 BY SIMILARITY.
FT CONFLICT 312 312 Y -> H (IN REF. 2).
SQ SEQUENCE 346 AA; 38178 MW; 8DAF887CAD1E3C1F CRC64;

```

```

Query Match 11.9%; Score 206; DB 1; Length 346;
Best Local Similarity 25.7%; Pred. No. 2.8e-11;
Matches 86; Conservative 47; Mismatches 117; Indels 84; Gaps 14;

OY 27 TORPRRTMSSTWELRPLVAVCVSTPNKQENFKCAELVQGAARLACALAPLPEF-- 84
DB 23 TVRTVTVOSTVYN-----DTPATIDAEKY-----IVPAASKAEVLVPEGEIG 69
OY 85 -----DFTARNPAETLLSEPLNGDLLGOYSOLARECGIMSLIG 124
DB 70 GYPRGFRFLGAVGHNEEGDEFRKYHASAI-----HYPGGEVARLADVARKNNVYLWGA 125
OY 125 FHERGQDWEQNOQIVQCHVLLNS-KGSVVASYKTHLCQVE--IPGGPMRESNYTKPGG 181
DB 126 IEKEG-----YLYL-CYVLFEPGQCFLOKHKRLMPTSLERICWGG-----DG 168
OY 182 TLEPPYKTRAGKVGALICVYMRPELSIKLAQGAELITPSPAFSGYSTGPAHMEVLLRAR 241
DB 169 STIPVYDTPIGKIGAAICWENRMPLYRTALYARGIELYCAPADGS---KEQSSMLHT 224
OY 242 AHSQCVVIAAO-CGR-----HHETRASYGSHSWYDPMGTVVARCS 282
DB 225 AITGGCFVLACQFCQKRPDPDLFTWYDDEKEDSIVSOGGSYIISPLQVLAGPN 264
OY 283 -EGPGCLARIDLHFLQQRHQL-PVFGHRRPDL 314
DB 285 FESEGLVTDADIGDIAARAKLYFDSYGVYSPDV 318

RESULT 12
NRL3_ARATH STANDARD; PRT; 346 AA.
AC P46010; O04909;
ID 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-NOV-2001 (Rel. 40, Last annotation update)
DE NITRILASE 3 (EC 3.5.5.1).
GN NIT3 OR AT3G44320 OR T10D17.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RN SEQUENCE FROM N.A.
RN STRAIN=CV, COLUMBIA;
RX MEDLINE=94294436; PubMed=8022831;
RA Bartel B., Fink G.R.;
RT "Differential regulation of an auxin-producing nitrilase gene family
RT in Arabidopsis thaliana."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).
RN [2]
RN SEQUENCE FROM N.A.
RN STRAIN=CV, COLUMBIA;
RX MEDLINE=98145459; PubMed=9484465;
RA Hillebrand H., Bartling D., Weiler E.W.;
RT "Structural analysis of the nitr/nitr/nit3 gene cluster encoding
RT nitrilases, enzymes catalyzing the terminal activation step in indole-
RT acetic acid biosynthesis in Arabidopsis thaliana."
RL Plant Mol. Biol. 36:89-99(1998).
RN [3]
RN SEQUENCE FROM N.A.
RN STRAIN=CV, COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unsel'd M.,
RA Fattmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Deleney M., Boutry M., Griwell L.A., Maché R., Pulgomech P.,
RA De Simone V., Choise N., Artiguenave F., Robert C., Brotier P.,
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Mumbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,

```

RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordiek G.,
RA Reichelt J., Scherfe M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,
RA Cooke R., Landie M., Berger-Liauro C., Furnelle B., Masuy D.,
RA de Haan M., Maestre A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argitlou A., Flores M., Lignori R., Vitale D.,
RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Meves H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T.H., Rizzo M., Malt R., Malt R., Peterson J., Van Aken S.,
RA Cressy T.H., Haas B., Malt R., Malt R., Peterson J., Van Aken S.,
RA Pal G., Miltischer J., Sellers P., Gill J.E., Feldlyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsuno M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
RA Matanabe A., Yamada M., Yasuda M., Tabata S.,
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:820-822(2000).
CC -I- FUNCTION: CAN CONVERT INDOL-3-ACETONITRILE TO THE PLANT HORMONE
CC INDOL-3-ACETIC ACID.
CC -I- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
CC -I- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.
CC -I- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; U09959; AAA19627.1; -;
DR EMBL; Y07648; CA68936.2; -;
DR EMBL; AL338865; CAB89000.1; -;
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR000132; Nitril_cyn_hydrolase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
DR HydroLase; MultiGene family.
RT ACT SITE 186 186 BY SIMILARITY.
SQ SEQUENCE 346 AA; 38022 MW; 70CFF421547F2B5E CRC64;

Query Match 11.7%; Score 203; DB 1; Length 346;
Best Local Similarity 24.7%; Pred. No. 5.3e-11;
Matches 82; Conservative 50; Mismatches 116; Indels 84; Gaps 13;

QY 30 RPRMTSSSTSWELPLVAVQVSTPNKQENFTCAELVQEARLACLAFLPEAF----- 84
DB 26 RVTIVOSTTVN-----DTPATLDKAEK-----IVEASKGAKVLDFPEAFEGV 72
QY 85 -----DFIARNPAETLLSEPLNGDLGOYOLARCGIMLSLGGPH 126
DB 73 RGRFGLAVGVNIEGRDF--RNYHASNI--KVPGEVELELAELAGNNVHLVGALE 127
QY 127 ERGDMEONKITYNCVHLNKGVSVAASYRKTILCDVE--IPGQPMRESNVTYKPGTLE 184
DB 128 KDG-----YTLVCTALFFSPQGFGLKHKRYMPTSLERCINGOG-----DGSIT 171
QY 185 PPVTPACKVGLATCYDMARFPELSIKLAQACAEILTPSAFSSVTPGPAWEVLLARALE 244
DB 172 PVTDPICKIGALICWENRMPLYRTALYAKGIEICAPLADSL-----EMQASMIHIAE 227
QY 245 SOCYVIAAO-CGR-----HHEFRASVGSVMVDPGTVAAACS-BG 284
DB 228 GGCFFVLSHQCRKREFPEHPDYLFENDIVDTKEHPVTVSGGGSVITISPLGKVLADPNYES 287

QY 285 PGLICARIDLHFLQONKHLVPFOH-RRPDLY 315
DB 288 ESLVTAADLDGIARAKLYFDVGHYSKPDIF 319
RESULT 13
BUP_RAT
ID BUP_RAT STANDARD: PRT; 393 AA.
AC 003248;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BUP-UREIDOPROPIONASE (EC 3.5.1.6) (BETA-ALANINE SYNTHASE)
DE (N-CARBAMOYL-BETA-ALANINE AMIDOHYDROLASE).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 202-212.
RC TISSUE=Liver;
RX MEDLINE=93194066; PubMed=8449931;
RA Kvalnes-Krick K.L., Traut T.W.;
RT "Cloning, sequencing, and expression of a cDNA encoding beta-alanine
RT synthase from rat liver.";
RL J. Biol. Chem. 268:5686-5693(1993).
CC -I- CATALYTIC ACTIVITY: N-CARBAMOYL-BETA-ALANINE + H(2)O =
CC BETA-ALANINE + CO(2) + NH(3).
CC -I- COFACTOR: BINDS TWO ZINC IONS.
CC -I- ENZYME REGULATION: ALLOSTERIC ENZYME WITH POSITIVE COOPERATIVITY
CC TOWARD THE SUBSTRATE N-CARBAMOYL-BETA-ALANINE.
CC -I- PATHWAY: THIRD AND FINAL STEP IN THE CATABOLISM OF THE
CC PYRIMIDINE BASES, URACIL OR THYMINE, TO PRODUCE BETA-ALANINE OR
CC 2-METHYL-BETA-ALANINE.
CC -I- SUBUNIT: IN THE ABSENCE OF LIGANDS, THE ENZYME EXISTS AS A STABLE
CC HOMODIMER, ALTHOUGH THIS IS NOT THE MOST ACTIVE FORM OF THE
CC ENZYME. THIS NATIVE HEXAMER DISSOCIATES TO AN INACTIVE TRIMER IN
CC RESPONSE TO THE PRODUCT, BETA-ALANINE, OR ASSOCIATES TO THE MORE
CC ACTIVE HOMODIMER IN RESPONSE TO THE SUBSTRATE.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; M97662; AAA40804.1; -;
DR PIR; S27881; S27881.
DR PIR; A46624; A46624.
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam; PF00795; CN_hydrolase; 1.
KW HydroLase; Zinc.
FT MOD_RES 71 71 BLOCKED.
FT METAL 97 97 ZINC 1 (POTENTIAL).
FT METAL 101 101 ZINC 1 (POTENTIAL).
FT METAL 158 158 ZINC 1 (POTENTIAL).
FT METAL 280 280 ZINC 2 (POTENTIAL).
FT METAL 293 293 ZINC 2 (POTENTIAL).
FT METAL 297 297 ZINC 2 (POTENTIAL).
SQ SEQUENCE 393 AA; 44042 MW; 723E5AF5B01E3AAD CRC64;

Query Match 11.1%; Score 193; DB 1; Length 393;
Best Local Similarity 24.7%; Pred. No. 5e-10;
Matches 71; Conservative 40; Mismatches 147; Indels 30; Gaps 6;

QY 51 TSPNKE--NFKCALVEQEARLGLATFLPEA---PFIIARNPAETLLSEPLNG 103
DB 85 TSAPVABOVSALHKRIETAEVAAACGVNITCFQEAAMMPFACTREKLPTEFASAD 144

OY 104 DLGGYSQIARECGIWLISGFGHERGQDWEONKRIYCNHVLNKSQSVVAYSRKTLCDV 163
 DB 145 GILTRRCQKLAHKNHNVVSPILERRD--HGVLYMNTAVVINSGLVWCKTKRKNH---- 198
 OY 164 ETPGQPMRESNRYKGGTLEPPVKTPAGKVGAIICYDMRPELSIKLAOGAEILTYPS 223
 DB 199 -IPRVQDFNESTYVMGNGHGFQFQFGRIVANVICYGRHHPHMLMYVNGAEIIFNPS 257
 OY 224 AGSGVYGRHWEVLRARALIESOCYIAAAGCR-----HHETRASYGHS 268
 DB 258 ATGELSESMWPLEARNAAIAHNCFCALNRVGOEHYPNEFTSGDKKAAHDLGYEYSS 317
 OY 269 MYVDPGTVVARSCEG-PEGLTARIDLHFLQGRHLPVQHRPPDY 315
 DB 318 YVAAPDGRTPGLSRQDGLVLTENLNCQOINDPWTFRKMTGRLEMY 365
 RESULT 14
 NR12_ARATH STANDARD: PRT: 339 AA.
 AC P32962; Q96505; (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NITRILASE 2 (EC 3.5.5.1).
 GN NIT2 OR AT3G44300 OR T10D17_90.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERECTA; TISSUE=Leaf;
 RX MEDLINE=94286570; PubMed=8016109;
 RA Bartling D., Seedorf M., Schmidt R.C., Weiler E.W.;
 RT "Molecular characterization of two cloned nitrilases from Arabidopsis
 thaliana: key enzymes in biosynthesis of the plant hormone indole-3-
 acetic acid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=94294436; PubMed=8022831;
 RA Bartel B., Fink G.R.;
 RT "Differential regulation of an auxin-producing nitrilase gene family
 in Arabidopsis thaliana.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Zhou L., Bartel B., Thornburg R.W.;
 RT "Nucleotide sequence of a pathogen induced nitrilase gene from
 Arabidopsis thaliana: Nlt2.";
 RL (In) Plant Gene Register PGR96-006.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Farmanbat M., Lemcke K., Rieger M., Ansoorge W., Unseld M.,
 RA Sattmann M., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
 RA Deleseny M., Boutry M., Griwall L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choisme N., Artiguenave F., Robert C., Brotlier P.,
 RA Winkler P., Catolico L., Weissenbach J., Saurin W., Queller F.,
 RA Schuchert M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurnbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
 RA Wieselmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.H., Nordsiek G.,
 RA Reichelt J., Scharte M., Schoen O., Barges M., Terol J., Clement J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laurie M., Berger-Liauro C., Purnelle B., Masuy D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,

RA Monfort A., Argitrou A., Flores M., Liguori R., Vitale D.,
 RA Manhaupt G., Haase D., Schoof H., Rüd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
 RA Preuss D., Lin X., Niernm W.C., Salzberg S.L., White O., Venter J.C.,
 RA Frazer C.M., Kaneko T., Nakamura Y., Sato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Ideasa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Kiyokawa S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Matanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:820-822(2000).
 CC -1- FUNCTION: CAN CONVERT INDOL-3-ACETONITRILE TO THE PLANT HORMONE
 INDOL-3-ACETIC ACID.
 CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
 CC -1- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.
 CC -1- DEVELOPMENTAL STAGE: BARELY DETECTABLE IN YOUNG ROSETTES, BUT IS
 STRONGLY EXPRESSED DURING BOLTING, FLOWERING, AND ESPECIALLY FRUIT
 DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X68305; CAA48377.1; -;
 CC EMBL: U09958; AAB60275.1; -;
 CC EMBL: U38845; AAB05220.1; -;
 CC EMBL: AL353865; CAB8998.1; -;
 CC PIR: S31969; S31969.
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR000132; Nitril_cyn_hydrolase.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS00920; NITRIL_CHT_1; 1.
 DR PROSITE: PS00921; NITRIL_CHT_2; 1.
 KW Hydrolase; Multigene family.
 FT ACT_SITE 179 179 BY SIMILARITY.
 FT CONFLICT 37 37 E -> G (IN REF. 3).
 FT CONFLICT 48 48 S -> T (IN REF. 3).
 SQ SEQUENCE 339 AA; 37153 MW; 06CDE76D2FDC24A7 CRC64;
 Query Match 10.88; Score 187.5; DB 1; Length 339;
 Best Local Similarity 24.28; Pred. No. 1.3e-09;
 Matches 79; Conservative 51; Mismatches 124; Indels 73; Gaps 12;
 OY 34 MSSSTWELPVAACVY-TSPNKGNEKTCALVEOARAGACLAFLPEAF----- 84
 DB 12 VASSTYRATVQASTYVNDTPATLEK--ANKTIVBASGSLVVPPEAFIGYRGF 68
 OY 85 -----DFTARNPAETLLSLEPLNDLGOYSOLARECGIWLISGFGHERGQ 130
 DB 69 RFLGVGVHNEDEGRDEPRKYNASAI-----KVPGEVEKLAEKANNVYLWGAIEKGG- 123
 OY 131 DWEDNOKIYCNHVLNKSQSVVASYRKTILCDVE--IPGQPMRESNRYKGGTLEPPVK 188
 DB 124 -----YLYCTAFLFFSQGGFLGKRRMLPTSLRRCITWGG-----DGSTIPYD 168
 OY 189 TPAGKVGAIICYDMRPELSIKLAOGAEILTYPSAGSVYGRHWEVLRARALIESOCY 248
 DB 169 TPAGKVGAIICYDMRPELSIKLAOGAEILTYPSAGSVYGRHWEVLRARALIESOCY 224
 OY 249 VVAAQ-C-----GRHETRASYGHSNVDWGVVARS-EGPGLC 288
 DB 225 VLSACQCLRKDPFDHDPVLTDMYDDEKPSIVSOGSGVYIISPLGVLAGPNESEGCLI 284

OY 289 TARDLHFLQOMRQHL-PVFOHRRPDL 314
 DB 285 TADLDGDVAKLYEDSVGHYSRPDV 311

RESULT 15

ALAM_PSEAE STANDARD: PRT; 346 AA.

ID ALAM_PSEAE

AC P11436;

DT 01-OCT-1989 (Rel. 12, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DE ALIPHATIC AMIDASE (EC 3.5.1.4).

GN AMIE OR PA3366.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

CC Pseudomonas.

OX NCBI_TaxID=287;

RP [1]

RC STRAIN=PA3142;

RX MEDLINE=87219101; PubMed=3108029;

RA Ambler R.P., Auffret A.D., Clarke P.H.;

RT "The amino acid sequence of the aliphatic amidase from Pseudomonas aeruginosa."

RL FEBS Lett. 215:285-290(1987).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=87219102; PubMed=3108030;

RA Brammar W.J., Charles I.G., Matfield M., Liu C.-P., Drew R.E.,

RA Clarke P.H.;

RT "The nucleotide sequence of the amle gene of Pseudomonas aeruginosa."

RL FEBS Lett. 215:291-294(1987).

RN [3]

RP SEQUENCE FROM N.A.

RX STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen."

RL Nature 406:959-964(2000).

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen."

RL Nature 406:959-964(2000).

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

Query Match 10.7%; Score 186.5; DB 1; Length 346;
 Best Local Similarity 25.4%; Pred. No. 1.7e-09;
 Matches 73; Conservative 44; Mismatches 113; Indels 57; Gaps 14;

OY 56 KOENFTCAELVQEARL-----GACLAFLPE-AEDFARNPAETILLSEPTL 102
 DB 21 KMERLHTAAEVLDMARKIAEMIVGMKQGLPMDLVFPEYSLQIGIMYPDAEMMETAVAI 80
 OY 103 GDLGQYOLARRCIGIWLISLGFHERGODMEONOK--IYNCHVILNSKGSVASYRK-TH 159
 DB 81 GEETELFSRACKRANW---GVPSLTGERHEHPRAPNTLVLDINNEIYQYKRIIP 137
 OY 160 LCDVE--IPGQPMRESNTKPGCTLEPPVKTAPG-KVGLAICYDMRFPFLSLKLAQAGA 216
 DB 138 WCPLEGWYPG-----GOTVYSEG-----PKGKISLILICDDGNYPEIWRDCAMKGA 183
 OY 217 EILT-----YSAGSYVGRPAHVEYLARALESCYVIAAOCGRHETASYSHSW 270
 DB 184 ELIVRCGYMYPKADQV-----MMARAMANNQYV-AVANAAGPDGVSYSFGSHAI 235
 OY 271 VDPWGTIVARCEGP-GLCLARIDLHFLQ-----OMROHLPVFOHR 310
 DB 236 IGFDTGLGECGEEMGIYQAQLSLQIRDAANDQSQNHLEKILHR 282

Search completed: April 26, 2002, 17:02:50
 Job time: 453 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:22 ; Search time 57.61 Seconds

(without alignments)
820.100 Million cell updates/sec

Title: US-09-357-675C-22
Perfect score: 1736
Sequence: 1 MLEGTTRPPHQLCTGYRL.....LPVQHRRPDLXSLGHPIS 323

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SPTEMBL_17:*
2: sp.archaea:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.invertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1736	100.0	323	11	088526
2	1731	99.7	323	11	088526
3	1554	89.5	290	11	092BF2
4	1487	85.7	327	4	076091
5	921	53.1	288	13	0918G0
6	706	40.7	440	5	076463
7	699.5	40.3	460	5	076464
8	632.5	36.4	316	10	091E50
9	588.5	33.9	276	3	094660
10	538.5	31.0	272	1	027839
11	491	28.3	270	2	083040
12	483	27.8	276	4	09NOR4
13	482	27.8	282	2	09HYU6
14	470.5	27.1	276	11	09JHW2
15	431.5	24.9	275	2	09KUD4
16	427.5	24.6	283	2	09A9Y3
17	417	24.0	318	10	09LYH1
18	410	23.6	270	2	09KOW5
19	404	23.3	270	2	09JTL3

20	393.5	22.7	189	11	09CTG9	09ctg9 mus musc
21	389.5	22.4	283	5	09VHE4	09vhe4 drosophila
22	366	21.1	231	11	09D0S1	09d0s1 mus musc
23	342.5	19.7	113	6	029367	029367 sus scrofa
24	341.5	19.7	264	2	09XA70	09xa70 streptomyce
25	340	19.6	257	1	030121	030121 archaeoglob
26	334	18.5	297	2	09RRF7	09rrf7 deinococcus
27	332	18.5	270	1	09HTM8	09htm8 thermoplas
28	290.5	16.7	300	10	09XG19	09xg19 lycopersico
29	285.5	16.4	292	2	0916J8	0916j8 pseudomonas
30	276.5	15.9	259	2	031664	031664 bacillus su
31	276.5	15.9	271	2	09KE11	09ke11 bacillus ha
32	275	15.8	287	1	09Y9L1	09y9l1 aeropyrum p
33	271	15.6	270	2	091241	091241 pseudomonas
34	262.5	15.1	220	10	09ZOH4	09zoh4 arbidopsis
35	262.5	15.1	262	2	09L104	09l104 streptomyce
36	262.5	15.1	280	2	069939	069939 streptomyce
37	262.5	15.1	295	2	09PAQ2	09paq2 xyella fas
38	261.5	15.1	292	2	09ABL5	09abl5 caulobacter
39	257.5	14.8	282	2	09Z5U5	09z5u5 zymomonas m
40	255.5	14.7	280	2	069808	069808 streptomyce
41	255	14.7	298	12	089413	089413 paramacium
42	251.5	14.5	292	2	025079	025079 helicobacte
43	250.5	14.4	294	2	09ZMC7	09zmc7 helicobacte
44	249.5	14.4	261	2	099SL3	099sl3 staphylococ
45	248	14.3	292	2	025452	025452 helicobacte

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	323 AA.
088526	088526			
AC	088526			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	NITRILASE HOMOLOG 1.			
GN	NITL.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98337986; PubMed=9671749;			
RA	Pekarsky Y., Campigallo M., Siprashvili Z., Druck T., Sedkov Y.,			
RA	Tilib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,			
RA	Buchberg A.M., Maro A., Brenner C., Croce C.M.,			
RT	"Nitrilase and Fhl homologs are encoded as fusion proteins in			
RT	Drosophila melanogaster and Caenorhabditis elegans."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).			
DR	EMBL: AF069985; AAC40184.1; -			
DR	MGI: MGI:1350916; Ntl.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	InterPro: IPR001110; UPF0012.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
SO	SOURCE 323 AA; 35717 MW; 10151CER15IDFZC7 CRC64;			

Query Match 100.0%; Score 1736; DB 11; Length 323;
Best Local Similarity 100.0%; Pred. No. 5.3e-151;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLEGTTRPPHQLCTGYRLRIPVCTQPRRTMSSSTWELPYAACVTPNKGNF 60
DB	1	MLEGTTRPPHQLCTGYRLRIPVCTQPRRTMSSSTWELPYAACVTPNKGNF 60
QY	61	KTCAELVQEARLGCACAFPEAFDFIARNPAETILLSEPLNGDLGQYSQLARECGIWL 120
DB	61	KTCAELVQEARLGCACAFPEAFDFIARNPAETILLSEPLNGDLGQYSQLARECGIWL 120

OY 121 SLGGFHERGQDMEQOKIYNCHVLLNSKGSVYASRYKTHLCVETIPGQPMRESNNTKRG 180
DB 121 SLGGFHERGQDMEQOKIYNCHVLLNSKGSVYASRYKTHLCVETIPGQPMRESNNTKRG 180
OY 181 GTLEPPVKTGAGVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240
DB 181 GTLEPPVKTGAGVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240
OY 241 RAIESQCVYIAAOCGRHETRASVGHSMVDPKGTIVVARGSEGGCLARIDLHFLQOM 300
DB 241 RAIESQCVYIAAOCGRHETRASVGHSMVDPKGTIVVARGSEGGCLARIDLHFLQOM 300
OY 301 ROHLPVFOHRRPDLVGSGLHPLS 323
DB 301 ROHLPVFOHRRPDLVGSGLHPLS 323

RESULT 2
O9R1N4 PRELIMINARY: PRT: 323 AA.
AC O9R1N4.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98337986; PubMed=9671749;
RA Pekarsky Y., Campigillo M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganescu A., Wernuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and flit homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL: AF069988; AAC40185.1; -.
DR MGD: MGI:1350916; NtL1.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase: 1.
DR PROSITE: PS01227; UPF0012: 1.
SQ SEQUENCE 323 AA: 35705 MW: F8CD7730713665EF CRC64;

Query Match 99.7%; Score 1731; DB 11; Length 323;

Best Local Similarity 99.7%; Pred. No. 1.5e-150;

Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MIGETTRPHOLLCTGYRLRLIPVLCOTOPRPTMSSSTSWELPLVAVCOVSTPNKQENF 60
DB 1 MIGETTRPHOLLCTGYRLRLIPVLCOTOPRPTMSSSTSWELPLVAVCOVSTPNKQENF 60
OY 61 KTCALVNGEAAALGACLAFLPEAFDFIARNPATETLLSEPLNGDLLGOYSOLAREGIVL 120
DB 61 KTCALVNGEAAALGACLAFLPEAFDFIARNPATETLLSEPLNGDLLGOYSOLAREGIVL 120
OY 121 SLGGFHERGQDMEQOKIYNCHVLLNSKGSVYASRYKTHLCVETIPGQPMRESNNTKRG 180
DB 121 SLGGFHERGQDMEQOKIYNCHVLLNSKGSVYASRYKTHLCVETIPGQPMRESNNTKRG 180
OY 181 GTLEPPVKTGAGVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240
DB 181 GTLEPPVKTGAGVGLAICYDMRPFELSLKLAQAGAEILTPSARGSVTPGAHWEVLLRA 240
OY 241 RAIESQCVYIAAOCGRHETRASVGHSMVDPKGTIVVARGSEGGCLARIDLHFLQOM 300
DB 241 RAIESQCVYIAAOCGRHETRASVGHSMVDPKGTIVVARGSEGGCLARIDLHFLQOM 300

OY 301 ROHLPVFOHRRPDLVGSGLHPLS 323
DB 301 ROHLPVFOHRRPDLVGSGLHPLS 323

RESULT 3
O9DBF2 PRELIMINARY: PRT: 290 AA.
AC O9DBF2.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glassl C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schiml L.M., Stabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004988; BAB33723.1; -.
DR MGD: MGI:1350916; NtL1.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase: 1.
DR PROSITE: PS01227; UPF0012: 1.
SQ SEQUENCE 290 AA: 31886 MW: 34493DBFE2170C71 CRC64;

Query Match 89.5%; Score 1554; DB 11; Length 290;

Best Local Similarity 100.0%; Pred. No. 2.2e-114;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 MSSSTSMELPLVAVCOVSTPNKQENFKAELVDEAALGACLAFLPEAFDFIARNP 93
DB 34 MSSSTSMELPLVAVCOVSTPNKQENFKAELVDEAALGACLAFLPEAFDFIARNP 93
OY 94 TLLSEPLNGDLLGOYSOLAREGIVLSIGFHERGQDMEQOKIYNCHVLLNSKGSVYA 153
DB 94 TLLSEPLNGDLLGOYSOLAREGIVLSIGFHERGQDMEQOKIYNCHVLLNSKGSVYA 153
OY 154 STRKTHLCVETIPGQPMRESNNTKRGTLPEPVTPAGVGLAICYDMRPFELSLKLAQ 213
DB 154 STRKTHLCVETIPGQPMRESNNTKRGTLPEPVTPAGVGLAICYDMRPFELSLKLAQ 213
OY 214 AGAELITPSARGSVTPGAHWEVLLRARIESQCVYIAAOCGRHETRASVGHSMVDP 273
DB 214 AGAELITPSARGSVTPGAHWEVLLRARIESQCVYIAAOCGRHETRASVGHSMVDP 273
OY 274 WGTIVVARGSEGGCLARIDLHFLQOMROHLPVFOHRRPDLVGSGLHPLS 323
DB 274 WGTIVVARGSEGGCLARIDLHFLQOMROHLPVFOHRRPDLVGSGLHPLS 323


```
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL Science 282:2012-2018(1998).
DR EMBL: AF069986; AAC39136.1; -
DR EMBL: AL132860; CAB60517.1; -
DR HSSP: P49789; 2FTT.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001310; HIT.
DR InterPro: IPR001110; UPE0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR Pfam: PF01230; HIT; 1.
DR PROSITE: PS00892; HIT; 1.
DR PROSITE: PS01227; UPE0012; 1.
SQ SEQUENCE 440 AA; 49936 MW; 34EED23CEPF2975 CRC64;

Query Match 40.7%; Score 706; DB 5; Length 440;
Best Local Similarity 48.6%; Pred. No. 1.7e-56;
Matches 138; Conservative 51; Mismatches 89; Indels 6; Gaps 2;

OY 32 RTMSSTSELPVAVCVSTSPNKOENFKTCAELVQEARIGACIAPLEAFPIANRP 91
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 8 RTMATGRHF---IAVCOMTSNDLEKNFOAKNMIEERAGEKKCEWVLECEDEFGLNK 63
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 32 AETLLSEPLNDLGGYSOLARECGIWLISGFGHERGQDMQNKIYNCHVLLNSKGSV 151
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 64 NEOIDLAAATDCEYMEKRELARKNINWISLGGLNHK--DPSDAHPWMTLLITSDGVT 121
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 152 VASVYKTHLCDEYVIGPGPMRESNYTKPGGLEPPVKTGKGVGLAICYDMRPELSLKL 211
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 122 RAEYKMLHFDLEIGKVLMESEFSKAGTEWIPVDPDPIGLGISICYDAVPELSLMN 181
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 212 AQAGAEILTPRASFVSGPRAHWEVLLRARATESOCYVIAAOCGRHHTRASVGHSMVY 271
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 182 RRRGAQLLSFPAFTLNLGIAHWETLLRRATEINOCYVAAAGTSAHNPKRSGSHSMVY 241
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 272 DPMGVTVARCSGPGICLARIIDHFLQOMROHLPVFOHRRPPLY 315
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 242 DPMGVAVVAGCSRVDMCFEAELIDSLVDTLRMQPVFSHRSDLY 285
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 7
ID 076464 PRELIMINARY; PRT: 460 AA.
AC 076464;
DT 01-NOV-1998 (TREMBlrel. 08. Created)
DT 01-NOV-1998 (TREMBlrel. 08. Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17. Last annotation update)
DE NINEHIT PROTEIN.
GN NINEHIT OR CG7067.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridoidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Sauton G.G., Rogers J., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers J., Yandell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beaskey E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
```

```
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbagwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paclebo J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98337986; PubMed=9671749;
RA Pedersen Y., Campigillo M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilius S., Draganesco A., Wernuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Pht homologs are encoded as fusion proteins in
RL Drosophila melanogaster and Caenorhabditis elegans."
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL: AE003467; AAF47347.1; -
DR EMBL: AF069989; AAC39137.1; -
DR HSSP: P49789; 2FTT.
DR FLYBase: FBgn0024945; NtPht.
DR InterPro: IPR001310; CN_hydrolase.
DR InterPro: IPR003010; HIT.
DR Pfam: PF00795; CN_hydrolase; 1.
DR Pfam: PF01230; HIT; 1.
DR PROSITE: PS00892; HIT; 1.
SQ SEQUENCE 460 AA; 52231 MW; 81121A0BC337706 CRC64;

Query Match 40.3%; Score 699.5; DB 5; Length 460;
Best Local Similarity 50.2%; Pred. No. 7.1e-56;
Matches 140; Conservative 44; Mismatches 88; Indels 7; Gaps 3;

OY 45 VAVCVSTSPNKOENFKTCAELVQEARIGACIAPLEAFPIANRPETLLSPPLNGD 104
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 35 IAVGMGRSTSDAAALSOVLELDVDAKSONCMFLPCCDCDVGSSRQTLELSGLDGE 94
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 105 ILGGYSOLARECGIWLISGFGHERGQDMQNKIYNCHVLLNSKGSVVASVYKTHLCDE 164
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 95 LMAQYRELAKCKKIMWISLGVYHER-----NDQKIFNAVILLNKGELAAVYRKLMPEYV 149
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 165 IPGGMRESNYTKPGGLEPPVKTGKGVGLAICYDMRPELSLKLAAQAEIITYPSA 224
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 150 TK-EVRLRESPTVTPGXYCLERVSYPVGQIGIQIDYDRFAEPALVLRKLANLTPYSA 208
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 225 FGSVVGPAHWEVLLRARATESOCYVIAAOCGRHHTRASVGHSMVYDPMGVTVARCSG 284
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 209 FTYAVGKAHWEVLLRARATEOCFVVAQAQIGMHNQKQSGHSMVSPMGVNLADCSQ 268
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

OY 285 P-GIACIARIIDHFLQOMROHLPVFOHRRPPLYGSLGHP 322
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 269 ELIDTAEDVLSVLSQSLQTMPCEFRHRNDIYALTAYNL 307
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 8
```


DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 KW Hydrolase; Complete Proteome.
 SQ SEQUENCE 272 AA; 30009 MW; EEB1A60EFABE7AB CRC64;

Query Match 31.0%; Score 538.5; DB 1; Length 272;
 Best Local Similarity 39.6%; Pred. No. 2,1e-41;
 Matches 109; Conservative 54; Mismatches 99; Indels 13; Gaps 4;

OY 45 VAVCOVSTPNKQENFKCAELVQEAARIGACLAFLPEAFDRIARPAETLLIS---EPL 101
 DB 3 LQICOMQVTEDEKKNVTAEMIVEAERGARLIYLPMEFTC---PYDNLFESEVAEDE 58
 OY 102 NDDLLGQYQLARECGIMWLSLGGFHERGQDMQONKIYCNHVLNSKGSVVASRKTHLC 161
 DB 59 NGYSIRTRMSIARELGILHIVAGSIPE---ETPEGIYNTSFVMDGNITGKRRVHLF 113
 OY 162 DVEIFGQGFMRSENYTKPGCTLEPPVKTGPKGVGLAICYDMRFPPELTLKLAQGAELITY 221
 DB 114 DINVGELISFRSDSLINDSV-TYIETPQCVMGVIGICIDMFPPELSRMALGAEVLIF 172
 OY 222 PSAFSGSVTPGAHWEVLLRARAIESQCYVIAAOCGRHETRASYSYGVVDPMGTIVARC 281
 DB 173 PGAFNMTGTPAHMRLIVRSALDNOCYVAASPARNPASAYAVGHSMIADPMGSVCEA 232
 OY 282 SEGPGLCLARIDLHFLQOMRQHLVPVQHRPDLXG 316
 DB 233 DASPSVITADIDLEAVERIRRLPLLRNRPDVG 267

RESULT 11
 O83040 PRELIMINARY; PRT; 270 AA.
 AC 083040;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ORF270 PROTEIN.
 OS Plectonema boryanum.
 OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
 OX NCBI_TaxID=1184;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DS;
 RA Matsumura T., Fujita Y., Hase T.;
 RL "Cyanobacterial ferredoxin gene."
 RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AB017194; BAA32602.1;
 DR InterPro: IPR001110; UPF0012.
 DR InterPro: IPR003010; CN_hydrolase.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 SQ SEQUENCE 270 AA; 29639 MW; 131C447C00FBE740 CRC64;

Query Match 28.3%; Score 491; DB 2; Length 270;
 Best Local Similarity 38.8%; Pred. No. 4.6e-37;
 Matches 111; Conservative 44; Mismatches 89; Indels 42; Gaps 6;

OY 44 LVAVCQVSTPNKQENFKCAELVQEAARIGACLAFLPEAFDRIARPAETLLISLEPL 103
 DB 5 LAAAVQMNLPDLOKNLAQAEELIEIAVRGAELIGLPENFSFL-----G 49
 OY 104 DLLGQYSQ-----LARECGIMWLSLGGF---ERGDQMEONQKIYCNHVLN 146
 DB 50 DEEGKIAQDELAIVSEKFLKTMARQFYVTLGGGYPVVERG-----KYNTFALLVA 102
 OY 147 SKGSVVASRKTHLCDELTPGCGPMRESNYTKPGCTLEPPV--KTIPAGKVGAICTDMF 204
 DB 103 PNGEELAREKXHLFDVNLPGDNTYRES-ATVAGTRVPRIHPSKDLGNIGLSVCYDVF 161

OY 205 PELSLKLAQGAELITYPSAFSGSVTPGAHWEVLLRARAIESQCYVIAAOCGRHETRAS 264
 DB 162 PELYRQMTQGAELVLEVPAAFPATHTGKHQVLLQARAIENTCYVIAAPQTRHNSLRQS 221
 OY 265 YGHSNVVDPMGTVVAVRCSEGPGLCLARIDLHFLQOMRQHLVPVQHR 310
 DB 222 HGHAMIDPMGVILADAGEMTGAIAETIAPRLDQVRROMPSLOHR 267

RESULT 12
 O9N0R4 PRELIMINARY; PRT; 276 AA.
 AC O9N0R4;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NIT PROTEIN 2 (CUA002).
 GN NIT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed-10959838;
 RA Pace H.C., Hodawadekar S.C., Draganescu A., Huang J., Bleganowski P.,
 RA Pekarsky Y., Croce C.M., Brenner C.;
 RA "Crystal structure of the worm Nitfhl1 Rosetta Stone protein reveals a
 RT Nit tetramer binding two Fhl1 dimers."
 RL Curr. Biol. 10:907-917(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ADRENOCORTICAL TUMOR SECRETED ACTH;
 RA Xu X., Yang Y., Gao G., Xiao H., Chen Z., Han Z.;
 RL Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF284574; AAF67103.1;
 DR EMBL: AF260334; AAG44665.1;
 DR InterPro: IPR003010; CN_hydrolase.
 DR Pfam: PF00795; CN_hydrolase; 1.
 SQ SEQUENCE 276 AA; 30608 MW; 32FA797601A419C6 CRC64;

Query Match 27.8%; Score 483; DB 4; Length 276;
 Best Local Similarity 38.2%; Pred. No. 2.5e-36;
 Matches 107; Conservative 54; Mismatches 105; Indels 14; Gaps 6;

OY 38 TSWELPLVAVCQVSTPNKQENFKCAELVQEAARIGACLAFLPEAFD--FIARPAETL 95
 DB 2 TEFRLALIDL-QISSI--KSDVYTRACSFIRPAARQAKIVSLPFCFNSPYGAKYFPE-- 56
 OY 96 LLSPEPLNGDLQOYSQLAARECGIMWLSLGGFHERGQDMQONKIYCNHVLNSKGSVVAS 155
 DB 57 -YAEKIPGESTOKLSVAECSIVYLIGSIPE----EDAGKLVNCAVFGPDGILLAKY 110
 OY 156 RTHLDCVEIIPQCGPMRESNYTKPGCTLEPPVKTGPKGVGLAICYDMRFPPELTLKLAQAG 215
 DB 111 RKIHFLDIDVPKRIIFQESKTLSPDPSF-TFDTPYCRGIGICIDMRRAELAQIYAQNG 169
 OY 216 AELITYPSAFSGSVTPGAHWEVLLRARAIESQCYVIAAOCGRHETRASYSYGVVDPMG 275
 DB 170 COLLYPGAFNLTTPPAHWEVLLQSRRAVNDQYVAVTASPARDKASAYAVGHSHVVPWG 229
 OY 276 TVVAVCSBGPGLCLARIDLHFLQOMRQHLVPVQHRPDLX 315
 DB 230 EVLAKAGTEBAIVYSDIDKLKLAETROOLPVFRKRSIDL 269

RESULT 13
 O9HV06 PRELIMINARY; PRT; 282 AA.
 AC O9HV06;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:54:36 ; Search time 56.94 seconds
(without alignments)
420.191 Million cell updates/sec

Title: US-09-357-675C-22

Perfect score: 1736

Sequence: 1 MLCGTRPPHQLCTGYRL.....LPVQRHRRPDLGSLGHPLS 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_1101.*

1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1482	85.4	464	21	AAV68739
2	747	43.0	224	21	AAV68739
3	733	42.2	153	22	AAV10877
4	483	27.8	276	22	AAV10877
5	313	18.0	159	21	AAV10877
6	295.5	17.0	266	22	AAV10877
7	288.5	16.6	299	21	AAV10877
8	273.5	15.8	252	22	AAV10877
9	273.5	15.8	252	22	AAV10877
10	273.5	15.8	252	22	AAV10877
11	262.5	15.1	220	21	AAV10877

12	262.5	15.1	220	21	AAV10877
13	237	13.7	264	22	AAV10877
14	210	12.1	227	18	AAV10877
15	205	11.8	340	21	AAV10877
16	205	11.8	346	21	AAV10877
17	205	11.8	358	21	AAV10877
18	204	11.7	303	15	AAV10877
19	203	11.7	303	15	AAV10877
20	202	11.6	303	15	AAV10877
21	201.5	11.6	303	15	AAV10877
22	201	11.6	303	15	AAV10877
23	200.5	11.5	303	15	AAV10877
24	200	11.5	303	15	AAV10877
25	200	11.5	303	15	AAV10877
26	200	11.5	303	15	AAV10877
27	200	11.5	303	15	AAV10877
28	200	11.5	303	15	AAV10877
29	200	11.5	303	15	AAV10877
30	199.5	11.5	303	15	AAV10877
31	199.5	11.5	303	15	AAV10877
32	199.5	11.5	303	15	AAV10877
33	199	11.5	303	15	AAV10877
34	199	11.5	303	15	AAV10877
35	199	11.5	303	15	AAV10877
36	199	11.5	303	15	AAV10877
37	199	11.5	303	15	AAV10877
38	199	11.5	303	15	AAV10877
39	199	11.5	303	15	AAV10877
40	199	11.5	303	15	AAV10877
41	198.5	11.4	303	15	AAV10877
42	198.5	11.4	303	15	AAV10877
43	198.5	11.4	303	15	AAV10877
44	198.5	11.4	303	15	AAV10877
45	198.5	11.4	303	22	AAV10877

ALIGNMENTS

RESULT 1	AAV68739	standard; Protein; 464 AA.
ID	AAV68739	standard; Protein; 464 AA.
XX	AAV68739	
AC	AAV68739	
XX	AAV68739	
DT	05-MAY-2000	(first entry)
XX	05-MAY-2000	(first entry)
DE	Amino acid sequence of the N1L protein.	
XX	Amino acid sequence of the N1L protein.	
KW	N1L gene; nitrilase; tumour suppressor gene; FRIT; chromosome 3p14.2;	
KW	FRAB; cancer; genome allele inactivation.	
XX	FRAB; cancer; genome allele inactivation.	
OS	Homo sapiens.	
OS	Mus sp.	
OS	Drosophila melanogaster.	
OS	Caenorhabditis elegans.	
XX	Caenorhabditis elegans.	
XX	Caenorhabditis elegans.	
FT	Key	Location/Qualifiers
FT	Misc-difference 6	/note= "unspecified amino acid encoded by TNT"
FT	Misc-difference 6	/note= "unspecified amino acid encoded by TNT"
PN	WO200003685-A2.	
XX	WO200003685-A2.	
PD	27-JAN-2000.	
XX	27-JAN-2000.	
XX	20-JUL-1999;	99WO-US16366.
PF	20-JUL-1999;	99WO-US16366.
XX	20-JUL-1999;	99WO-US16366.
XX	20-JUL-1999;	99WO-US16366.
PR	(UYJE-) UNTV JEFFERSON THOMAS.	
XX	(UYJE-) UNTV JEFFERSON THOMAS.	
PA	Croce CM;	
XX	Croce CM;	
XX	Croce CM;	

DR WPI: 2000-171195/15.
DR N-PSDB; AAZ46101.

PT Novel nitrlase homologs used as diagnostic and therapeutic reagents
PT for the detection and treatment of cancer -

XX PS Disclosure: Fig 6; 25pp; English.

XX The present sequence is encoded by the coding region of human, murine,
CC Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC human and mouse NIT1 genes are members of an uncharacterised
CC mammalian gene family with homology to bacterial and plant nitrlases.
CC The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC for fusion proteins in which the Fhit domain is fused with a Nit domain.
CC In mouse and humans, FHIT and NIT are encoded by two different genes,
CC localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC chromosomal fragile site FRA3B, is often altered in most common forms
CC of human cancer. The NIT1 protein overcomes the mutated inactivation
CC of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC and analogues of them, and antibodies are used as diagnostic and
CC therapeutic reagents for the detection and treatment of cancers.

XX SQ Sequence 464 AA;

Query Match 85.4%; Score 1482; DB 21; Length 464;
Best Local Similarity 84.4%; Pred. No. 9.1e-146;
Matches 275; Conservative 27; Mismatches 20; Indels 4; Gaps 2;

QY 2 LGFTTRPPHQ--LLCTGYRLRLIPVLTQPRPTMS--STSWMLPLVAVCQVSTPNKQ 57
DB 37 lgftrpphrflllcpqlrlpqlsvlcagpramasssccelpvavcqvstpdqk 96
QY 58 ENFKCALVQEAARLGAFLPEAFDFIARNPAETLLSEPLNGDLLGOYSQARREG 117
DB 97 qnfkcaelvreearlgaclafldpaetlhlseplngkllleeytqlarec 156
QY 118 IWLISGFHERGQDEQONKTYNCHVLLNSKGSVAVSYRKTLCQVEIFGQPMRESNT 177
DB 157 lwlisgfherngqdeqgcklyncnvllnsskavvatyrkthldcvelpbggpmcesnt 216
QY 178 KPGTLEPPVKTTPAGKVGIAICYDMRPBELSLKLAQAGAEILTYPSAFSGVTPAHWEVL 237
DB 217 mpqgleasvstpaqkltglavcydmrfelsialagaeelltypsafsgtqphwevl 276
QY 238 LRAARLESCVYIAAOCGRHETRASGHSNVDPMGVAVARCEGSEGLCARIDML 297
DB 277 lraarleqcyvvaagcgrhnekrasyghsnvdpwglvarcseglclardlmyl 336
QY 298 QOMRQHLVFOHRRPDLVSGLSPLS 323
DB 337 rqlrrhlrvfqhrrpdlvynlghpls 362

RESULT 2
AAB57054
ID AAB57054 standard; Protein: 224 AA.

XX AC AAB57054;

XX DT 13-MAR-2001 (first entry)

DE Human prostate cancer antigen protein sequence SEQ ID NO:1632.

KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.

XX Homo sapiens.

XX WO200055174-A1.

XX PD 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.

XX PI Rosen CA, Ruben SM;

XX DR WPI: 2000-587513/75.

XX N-PSDB; AAF16257.

PT prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -

XX Claim 11; Page 2090-2091; 2338pp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytosolic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotoxic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.

XX SQ Sequence 224 AA;

Query Match 43.0%; Score 747; DB 21; Length 224;
Best Local Similarity 82.1%; Pred. No. 1.4e-69;
Matches 142; Conservative 15; Mismatches 12; Indels 4; Gaps 2;

QY 1 MLGFTTRPPHQ--LLCTGYRLRLIPVLTQPRPTMS--STSWMLPLVAVCQVSTPNK 56
DB 26 mlgftrpphrflllcpqlrlpqlsvlcagpramasssccelpvavcqvstpdqk 85
QY 57 QENFKCALVQEAARLGAFLPEAFDFIARNPAETLLSEPLNGDLLGOYSQARREG 116
DB 86 qnfkcaelvreearlgaclafldpaetlhlseplngkllleeytqlarec 145
QY 117 IWLISGFHERGQDEQONKTYNCHVLLNSKGSVAVSYRKTLCQVEIFGQPMRESNT 169
DB 146 lwlisgfherngqdeqgcklyncnvllnsskavvatyrkthldcvelpbggpmcesnt 198

RESULT 3

AAI10877
ID AAI10877 standard; Protein: 153 AA.

XX AC AAI10877;

XX DT 14-MAY-1999 (first entry)

DE Amino acid sequence of a human secreted protein.

KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;

XX	prostate disease; asthma; osteoporosis; arthritis.
OS	Homo sapiens.
XX	
FH	Key
FT	Peptide
FT	Protein
FT	Location/Qualifiers
XX	1..46
XX	/note= "signal peptide"
XX	47..153
XX	/note= "secreted protein"
XX	
PN	MO9907891-A1.
PD	18-FEB-1999.
XX	
PF	04-AUG-1998;
XX	98WO-US16235.
XX	
PR	19-AUG-1997;
PR	97US-0056732.
PR	05-AUG-1997;
PR	97US-0054798.
PR	05-AUG-1997;
PR	97US-0054803.
PR	05-AUG-1997;
PR	97US-0054804.
PR	05-AUG-1997;
PR	97US-0054806.
PR	05-AUG-1997;
PR	97US-0054807.
PR	05-AUG-1997;
PR	97US-0054808.
PR	05-AUG-1997;
PR	97US-0054809.
PR	05-AUG-1997;
PR	97US-0055309.
PR	05-AUG-1997;
PR	97US-0055310.
PR	05-AUG-1997;
PR	97US-0055312.
PR	05-AUG-1997;
PR	97US-0055386.
PR	05-AUG-1997;
PR	97US-0055311.
PR	18-AUG-1997;
PR	97US-0055970.
PR	18-AUG-1997;
PR	97US-0055986.
PR	19-AUG-1997;
PR	97US-0056365.
PR	19-AUG-1997;
PR	97US-0056366.
PR	19-AUG-1997;
PR	97US-0056557.
PR	19-AUG-1997;
PR	97US-0056370.
PR	19-AUG-1997;
PR	97US-0056371.
PR	19-AUG-1997;
PR	97US-0056563.
PR	19-AUG-1997;
PR	97US-0056731.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI	Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
XX	
XX	WPI: 1999-167452/14.
DR	N-PSDB; AAX30398.
XX	
PT	New isolated human genes encoding secreted polypeptides - useful for
PT	diagnosis and treatment of pathological diseases
XX	
PS	Claim 3; Page 304; 331pp; English.
XX	
CC	The specification describes secreted proteins and their corresponding
CC	polynucleotides which are useful for preventing, treating or ameliorating
CC	medical conditions, e.g. by protein or gene therapy. Pathological
CC	conditions can also be diagnosed by determining the amount of the
CC	secreted polypeptides in a sample or by determining the presence of
CC	mutations in the polynucleotides. Specific uses are described for each
CC	of the products, based on which tissues they are most highly
CC	expressed in, and include developing products for the diagnosis or
CC	treatment of cancer, tumours, neurodegenerative disorders, developmental
CC	abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC	diseases of the immune system, autoimmune diseases, hepatic and renal
CC	disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC	and cognitive disorders, schizophrenia, cardiovascular disorders,
CC	prostate diseases, asthma, disorders involving osteoclasts such as
CC	osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC	thymus, digestive/endocrine disorders, infections and AIDS. The
CC	polypeptides are also useful for identifying their binding partners.
XX	
XX	Sequence 153 AA;

[illegible][illegible]

Oy		156	RKHLGDEIFRGOGPMRSNTKRGGLTEPEPVKPAKVGALICYDRPPELSIKUAQAG	215
Dd		111	khlldfdvdygklttfesksklspgsdfs- lftdpccrvglgicymrfeelaqiyagrq	169
Oy		216	AELITTPSAGFSWGTPAHWEVLLRARAIESOCYVIAAAGCGRHETRASYGSHSMVDDPWG	275
Dd		170	cqlilvygpafnlcttgnhhwllqsrsvadngvyatasparddkasyyavghstlvnpwg	229
Oy		276	TIVARSGSGPGCLCARIDLHFLQGMROHLPFCFHRRDDLY	315
Dd		230	evlakgtteaalyvsydltkxlaetirqjlpvtfrtkrsldly	269
	RESULT	5		
	AAB42967	ID	AAB42967 standard; Protein: 159 AA.	
AC		XX	AAB42967;	
XX		XX		
DT		XX	08-FEB-2001 (first entry)	
DE		XX	Human ORFX ORF2731 polypeptide sequence SFG ID NO:5462.	
KW		KW	Human: open reading frame: ORFX; detection; cytosstatic; hepatotropic;	
KM		KM	vulnerary; antipsoatic; antiaparkinsonian; nootropic; neuroprotective;	
KM		KM	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	
KM		KM	hypotensive; thrombolytic; coagulant; vasotropic; antidiabetic;	
KM		KM	immunostimulant; dematological; immunosuppressive; antiinflammatory;	
KM		KM	antiviral; antibacterial; antifungal; antirheumatic; antihypoid;	
KM		KM	antihaemetic; gene therapy; cancer; proliferative disorder; hypertension;	
KM		KM	neurodegenerative disorder; osteoarthritis; graft vs host disease;	
KM		KM	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	
KM		KM	cholesterol ester storage; systemic lupus erythematosus; infection;	
KM		KM	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	
KM		KM	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	
KM		KM	bone damage; cartilage damage; antiinflammatory disease; coagulation;	
KW		KW	thrombosis; contraceptive.	
OS		XX	Homo sapiens.	
XX		XX		
PN		XX	WO200058473-A2.	
PD		XX	05-OCT-2000.	
PF		XX	31-MAR-2000; 2000WO-USO08621.	
PR		PR	31-MAR-1999; 99US-0127607.	
PR		PR	02-APR-1999; 99US-0127636.	
PR		PR	05-APR-1999; 99US-0127728.	
PR		PR	30-MAR-2000; 2000US-0540763.	
PA		XX	(CURA-) CURAGEN CORP.	
PI		XX		
PI		XX	Shimkets RA, Leach M;	
DR		XX	WPI: 2000-602362/57.	
N-PSDB:		XX	AAC771176.	
CC		XX	Novel nucleic acids and peptides derived from open reading frame X,	
PT		XX	useful for treating e.g. cancers, proliferative disorders,	
PT		XX	neurodegenerative disorders and cardiovascular disease -	
PS		XX	Claim 11; Page 4643; 5507pp; English.	
AA		XX	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,	
CC		XX	which represent the human ORFX open reading frames 1 to 3161. The ORFX	
CC		XX	sequences have activities such as: cytostatic; hepatotropic; vulnerary;	
CC		XX	antispasmodic; antiparkinsonian; nootropic; neuroprotective;	
CC		XX	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;	
CC		XX	immunosstimulant; cardiant; thrombolytic; coagulant; vasotropic;	
CC		XX	antidiabetic; hypotensive; dermatological; immunosuppressive;	
CC		XX	antifibromatory; antibacterial; antiviral; antifungal; antirheumatic;	
CC		XX	antithyroid; and antihaemetic. The sequences can be used for determining	

CC	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The
CC	nucleic acids can be used to express ORFX proteins in gene therapy
CC	vectors. The proteins and nucleic acids may be used to treat cancers,
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC	allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
CC	nutritional haemoglobinuria, antiinflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
SQ	Sequence 159 AA;
OY	Query Match 18.0%; Score 313; DB 21; Length 159;
Dd	Best Local Similarity 45.7%, Pred. No. 1.6e-24;
	Matches 59; Conservative 25; Mismatches 45; Indels 0; Gaps 0;
OY	167 VKTPRGVGLAICYMRPELSLKLAQAGELITPSPAFGSVTGPANHEVLRAPIESQ 246
	::::: :: :: :: : :::
Dd	24 letatcvtgdlgtcydmrtlaelaqlyaqgcqllyvpgafnlittpghnellgrstravng 83
OY	247 CTVIAAACGGHHETRASVGHSWVDVPMGVYVARCSSEGRGLCLARIDHFLDQMFOHLPV 306
	: : : : :: : : : :
Dd	84 yvvataasparddkaaayawahgstvnvpgevLaakgteaalysqdldkklaelrtqjpv 143
OY	307 FOHRRPDL 315
	:::
Dd	144 frqrtsdly 152
RESULT	6
AAG92814	
ID	AAG92814 standard; Protein: 266 AA.
AC	AAG92814;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum protein fragment SEQ ID NO: 6568.
XX	
KM	Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX	organic acid synthesis.
OS	Corynebacterium glutamicum.
PN	EPI108790-A2.
PD	20-JUN-2001.
PF	18-DEC-2000; 2000EP-0127688.
PR	16-DEC-1999; 99JP-0377484.
PR	07-APR-2000; 2000JP-0159162.
PR	03-AUG-2000; 2000JP-0280988.
PA	(KYOW) KYOWA HAKKO KOGYO KK.
XX	
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
DR	WPI: 2001-376931/40.
DR	N-PSDB; AAH68033.
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying
PT	muation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
XX	
P5	Claim 17; SEQ ID NO: 6568; 246bp + Sequence Listing; English.
XX	

CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from Corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX Sequence 266 AA;

Query Match 17.0%; Score 295.5; DB 22; Length 266;
Best Local Similarity 32.1%; Pred. No. 2.3e-22;
Matches 89; Conservative 46; Mismatches 113; Indels 29; Gaps 10;

QY 45 VAVCQVSTPKQENFKCAELVQEARLGLAFLPER-FEFIRNPAETILLSEPLNG 103
DB 3 Iallqistnsdkmdfiallrdaekaaeggarvlvfppeatqsfygrldt--gaeeIdg 60
QY 104 DLLAGYSLARECGIMLSLGFH-----ERGDWENONKRYNCVHLNKGSGVASYRKT 158
DB 61 efstavrkladeIdvIvIagmftpadtvgry---ektlsrvnvtvlisag-Ihngynkl 116
QY 159 HLCQVEIDFGQGMRESNNTKPGGTLEPPYKTPAGVGLAICYDMRPPELSLKLAQAGAEI 218
DB 117 htyda-----fygredsvkpgdel-vvfevddikfyvatcydtdirfpegfkIdarngaqI 170
QY 219 LTYPSAFSGSVGPA---HWEVILLRARALESQCYVIAAOCGNHETR-----ASYGSHM 269
DB 171 IvvpIswgd--gpgkleqewIpratrIdstcIvIacqgarIpeelrderkpgtIghsh 228
QY 270 VDPMGTVVARSCEGPGCLARIDHLPFOOMRHLPV 306
DB 229 vtphgveIasagyepeIdIadIvsgIakIrealpV 265

RESULT 7
AAG09683
ID AAG09683 standard; Protein: 299 AA.
XX AAG09683;
AC AAG09683;
XX 17-OCT-2000 (first entry)
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 7710.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7710.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EPI033405-A2.
FN EPI033405-A2.
XX 06-SEP-2000.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PE 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

Accession	Species	Gene	Protein	Length	Score	DB	Length	Score	DB
19-JUL-1999	99US-0144334			100	16.6%	288.5	21	299	16.6%
19-JUL-1999	99US-0144335			100	16.6%	288.5	21	299	16.6%
20-JUL-1999	99US-0144332			100	16.6%	288.5	21	299	16.6%
20-JUL-1999	99US-0144884			100	16.6%	288.5	21	299	16.6%
21-JUL-1999	99US-0144814			100	16.6%	288.5	21	299	16.6%
21-JUL-1999	99US-0145086			100	16.6%	288.5	21	299	16.6%
21-JUL-1999	99US-0145088			100	16.6%	288.5	21	299	16.6%
22-JUL-1999	99US-0145085			100	16.6%	288.5	21	299	16.6%
22-JUL-1999	99US-0145087			100	16.6%	288.5	21	299	16.6%
22-JUL-1999	99US-0145089			100	16.6%	288.5	21	299	16.6%
23-JUL-1999	99US-0145145			100	16.6%	288.5	21	299	16.6%
23-JUL-1999	99US-0145218			100	16.6%	288.5	21	299	16.6%
23-JUL-1999	99US-0145224			100	16.6%	288.5	21	299	16.6%
26-JUL-1999	99US-0145376			100	16.6%	288.5	21	299	16.6%
27-JUL-1999	99US-0145913			100	16.6%	288.5	21	299	16.6%
27-JUL-1999	99US-0145918			100	16.6%	288.5	21	299	16.6%
28-JUL-1999	99US-0145951			100	16.6%	288.5	21	299	16.6%
02-AUG-1999	99US-0146386			100	16.6%	288.5	21	299	16.6%
02-AUG-1999	99US-0146388			100	16.6%	288.5	21	299	16.6%
02-AUG-1999	99US-0146389			100	16.6%	288.5	21	299	16.6%
03-AUG-1999	99US-0147038			100	16.6%	288.5	21	299	16.6%
04-AUG-1999	99US-0147204			100	16.6%	288.5	21	299	16.6%
04-AUG-1999	99US-0147302			100	16.6%	288.5	21	299	16.6%
05-AUG-1999	99US-0147312			100	16.6%	288.5	21	299	16.6%
05-AUG-1999	99US-0147260			100	16.6%	288.5	21	299	16.6%
06-AUG-1999	99US-0147303			100	16.6%	288.5	21	299	16.6%
06-AUG-1999	99US-0147416			100	16.6%	288.5	21	299	16.6%
09-AUG-1999	99US-0147493			100	16.6%	288.5	21	299	16.6%
09-AUG-1999	99US-0147935			100	16.6%	288.5	21	299	16.6%
10-AUG-1999	99US-0148171			100	16.6%	288.5	21	299	16.6%
11-AUG-1999	99US-0148319			100	16.6%	288.5	21	299	16.6%
12-AUG-1999	99US-0148341			100	16.6%	288.5	21	299	16.6%
13-AUG-1999	99US-0148565			100	16.6%	288.5	21	299	16.6%
13-AUG-1999	99US-0148684			100	16.6%	288.5	21	299	16.6%
16-AUG-1999	99US-0149358			100	16.6%	288.5	21	299	16.6%
17-AUG-1999	99US-0149175			100	16.6%	288.5	21	299	16.6%
18-AUG-1999	99US-0149426			100	16.6%	288.5	21	299	16.6%
20-AUG-1999	99US-0149722			100	16.6%	288.5	21		

PR 25-FEB-1999;	99US-0121825.
PR 05-MAR-1999;	99US-0123180.
PR 09-MAR-1999;	99US-0123548.
PR 23-MAR-1999;	99US-0125788.
PR 25-MAR-1999;	99US-0126264.
PR 29-MAR-1999;	99US-0126785.
PR 01-APR-1999;	99US-0127462.
PR 06-APR-1999;	99US-0128234.
PR 08-APR-1999;	99US-0128714.
PR 16-APR-1999;	99US-0128845.
PR 19-APR-1999;	99US-0130077.
PR 21-APR-1999;	99US-0130544.
PR 23-APR-1999;	99US-0130510.
PR 23-APR-1999;	99US-0130891.
PR 28-APR-1999;	99US-0131449.
PR 30-APR-1999;	99US-0132048.
PR 30-APR-1999;	99US-0132407.
PR 04-MAY-1999;	99US-0132484.
PR 05-MAY-1999;	99US-0132485.
PR 06-MAY-1999;	99US-0132486.
PR 06-MAY-1999;	99US-0132487.
PR 07-MAY-1999;	99US-0132866.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
<hr/>	
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 10-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.6%; Score 288.5; DB 21; Length 299;
Best Local Similarity 28.2%; Pred. No. 1.5e-21;
Matches 82; Conservative 56; Mismatches 122; Indels 31; Gaps 9;

QY 44 LVACQVSTPNKQNEFTCAELVQEARLQCLAFLEAPF---FIRNPAETLLSEP 100
Db 11 vvsslqfscsdstcnuvaaaerlveaakganlllqelltegyyfcqagredffktrakp 70
QY 101 L-NGDLQOYSOLARECCIMWLSLGFHERGQDWEONKRYNCVHLNSKGSVVASYRRTH 159
Db 71 yknphltaimqklakelvgvlpvsffe-----eantahynslaidedgtdlgjyrksh 124
QY 160 LCDVEIPQOGR MRESNTRKPGGTLIEPVKTPAGKVGALICYDMRFPPELSLKLQAGAEI 218
Db 125 lpd-----gpqygekfyfnpqdtgfkvfqtkfaklgyvaicwdqvfpeaaravmlqgaei 178
QY 219 LTVPSAFGS-----VTPRPAHWEVLLRARATESQCVIAAAGCGR-----HHERRAS-Y 265
Db 179 lfypalaipsepdqglidrdhmwrvmqghaganvvlvasnlgkellelethgspqdlity 238
QY 266 GHSWVVDPMGTVAIVAKSCS-GRGLCLARIDLHFLQOMROHLPVFOCHRRPDLX 315
Db 239 gtsfiagptglgtvaeaddksaevlvaqfdldmlkkskrqswgvyfirrpdlly 289

RESULT 9
AAB79700
ID AAB79700 standard; Protein; 252 AA.
XX
AC AAB79700;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:134.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00923.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031421.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;

PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.

PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155458.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145066.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158363.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.8%; Score 205; DB 21; Length 340;
Best Local Similarity 24.7%; Pred. No. 9, 2e-13;
Matches 82; Conservative 50; Mismatches 116; Indels 84; Gaps 13;

QY 30 RPRMTSSSTSMELPIVAVCVTSTPNKQENFKTCALIEYOEANRLGACLAFLPEAF----- 84
DB 20 rvtlvqstlvyn-----dtpatlnhaekf-----iveaaskgaktvlflpeaflyggyp 66
QY 85 -----DFIARNPAETTLISEPLNGDLGOYSOLARECGIWLISLGFH 126
DB 67 rgfrfglawvhneegrde--rnyhasal---kvgpxverlaealagknvhlvymgaie 121
QY 127 ERGQDMEQNKIRYNCHVLLNSKGSVVAStRKTHLCDE--IPGQGPRESNYTKPGGTLE 184
DB 122 kdq-----yllyctalfifpgqgflgkhrkmpstlerciwgg-----dgsstl 165
QY 185 PPKTPAGKVGIAICYDMRPELSKLQAGAEILTYPSAFSGSVTGPAMHEVLLRAAIE 244
DB 166 pyvdeprgkigaaiowenmptlyrtalyakgleyccaptadystl-----ewqasmihaie 221
QY 245 SOCYVIAAAG-CGR-----HHETRASYGSHMVDPMCTVVAKCS-EG 284
DB 222 ggcftvlsahqfckrrretrepndylfndivdtkehdprvsggsvlspjklvlagpyses 281
QY 285 PGICLARIDLHFLQOMRQHLPEVQH--RRPDLY 315
DB 282 egivtadldigdiarakiyfdvavghyskpdf 313

Search completed: April 26, 2002, 16:54:36
Job time: 234 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:55:11 ; Search time 28.88 Seconds
(without alignments)
251.682 Million cell updates/sec

Title: US-09-357-675C-22

Perfect score: 1736

Sequence: 1 MGFTRPPHQLCTGYRL.....LPVQRHRRPDYSGSLGHPLS 323

Scoring table:

BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents,AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	204	11.8	303	2	US-08-294-871A-26	Sequence 26, Appl
2	204	11.8	303	2	US-08-294-871A-64	Sequence 64, Appl
3	204	11.8	303	3	US-08-876-398A-26	Sequence 26, Appl
4	204	11.8	303	3	US-08-876-398A-64	Sequence 64, Appl
5	203	11.7	303	2	US-08-294-871A-40	Sequence 40, Appl
6	203	11.7	303	2	US-08-876-398A-40	Sequence 40, Appl
7	202	11.6	303	2	US-08-294-871A-6	Sequence 6, Appl
8	202	11.6	303	2	US-08-876-398A-6	Sequence 6, Appl
9	201.5	11.6	303	2	US-08-294-871A-44	Sequence 44, Appl
10	201.5	11.6	303	2	US-08-876-398A-44	Sequence 44, Appl
11	201	11.6	303	2	US-08-294-871A-42	Sequence 42, Appl
12	201	11.6	303	2	US-08-294-871A-62	Sequence 62, Appl
13	201	11.6	303	3	US-08-876-398A-42	Sequence 42, Appl
14	201	11.6	303	3	US-08-876-398A-62	Sequence 62, Appl
15	200.5	11.5	303	2	US-08-294-871A-34	Sequence 34, Appl
16	200.5	11.5	303	2	US-08-294-871A-60	Sequence 60, Appl
17	200.5	11.5	303	2	US-08-876-398A-34	Sequence 34, Appl
18	200.5	11.5	303	3	US-08-876-398A-60	Sequence 60, Appl
19	200	11.5	303	2	US-08-294-871A-8	Sequence 8, Appl
20	200	11.5	303	2	US-08-294-871A-10	Sequence 10, Appl
21	200	11.5	303	2	US-08-294-871A-12	Sequence 12, Appl
22	200	11.5	303	2	US-08-294-871A-22	Sequence 22, Appl
23	200	11.5	303	2	US-08-294-871A-34	Sequence 34, Appl
24	200	11.5	303	3	US-08-876-398A-8	Sequence 8, Appl
25	200	11.5	303	3	US-08-876-398A-10	Sequence 10, Appl
26	200	11.5	303	3	US-08-876-398A-12	Sequence 12, Appl
27	200	11.5	303	3	US-08-876-398A-22	Sequence 22, Appl

ALIGNMENTS

```
RESULT 1
US-08-294-871A-26
; Sequence 26, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294, 871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971, 758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917, 111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211, 641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
```

28	200	11.5	303	3	US-08-876-398A-54	Sequence 54, Appl
29	199.5	11.5	303	2	US-08-294-871A-46	Sequence 46, Appl
30	199.5	11.5	303	3	US-08-876-398A-46	Sequence 46, Appl
31	199	11.5	303	1	US-07-917-111-2	Sequence 2, Appl
32	199	11.5	303	1	US-07-917-111-3	Sequence 3, Appl
33	199	11.5	303	1	US-08-479-638-2	Sequence 3, Appl
34	199	11.5	303	1	US-08-479-638-3	Sequence 3, Appl
35	199	11.5	303	2	US-08-294-871A-2	Sequence 2, Appl
36	199	11.5	303	2	US-08-294-871A-14	Sequence 14, Appl
37	199	11.5	303	2	US-08-294-871A-16	Sequence 16, Appl
38	199	11.5	303	2	US-08-294-871A-24	Sequence 24, Appl
39	199	11.5	303	2	US-08-294-871A-28	Sequence 28, Appl
40	199	11.5	303	2	US-08-294-871A-30	Sequence 30, Appl
41	199	11.5	303	2	US-08-294-871A-32	Sequence 32, Appl
42	199	11.5	303	2	US-08-294-871A-52	Sequence 52, Appl
43	199	11.5	303	2	US-08-294-871A-66	Sequence 66, Appl
44	199	11.5	303	3	US-08-876-398A-2	Sequence 2, Appl
45	199	11.5	303	3	US-08-876-398A-14	Sequence 14, Appl

```

1      FILING DATE: 27-DEC-1990
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: JP 078840/1991
4      FILING DATE: 11-APR-1991
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: JP 140051/1991
7      FILING DATE: 12-JUN-1991
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: PCT/JP92/00739
10     FILING DATE: 10-JUN-1992
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: JP 212692/1992
13     FILING DATE: 10-AUG-1992
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: PCT/JP93/01101
16     FILING DATE: 05-AUG-1993
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: JP 340078/1992
19     FILING DATE: 21-DEC-1992
20     ATTORNEY/AGENT INFORMATION:
21     NAME: Wegner, Harold C.
22     REGISTRATION NUMBER: 25, 258
23     REFERENCE/DOCKET NUMBER: 74129/127/AOPB
24     TELECOMMUNICATION INFORMATION:
25     TELEPHONE: (202) 672-5300
26     TELEFAX: (202) 672-5399
27     TELEX: 904136
28     INFORMATION FOR SEQ ID NO: 26:
29     SEQUENCE CHARACTERISTICS:
30     LENGTH: 303 amino acids
31     TYPE: amino acid
32     TOPOLOGY: linear
33     MOLECULE TYPE: protein
34     OS-08-294-871A-26

```

Query Match	11.8%	Score 204	DB 2	Length 303
Best Local Similarity	26.5%	Pred. No. 8.4e-14		
Matches 76; Conservative	46;	Mismatches 113;	Indels 52;	Gaps 12

QY	65	ELVOEAAALGCLAFLEP-----	-AFDFLARNAPAFELLSEPLNDLGGQ	SOLAREC	116	
Db	29	DMILTKAASRGANFLVYFRRLATLTFFRRKHFDEDELDSDSFEYE	-MGRPVVRFLEKRAEL		87	
QY	117	GIWLSL-----	GGEHNGSDMEONOKLYNCVLLNSKGSVVA	RYKTHLDVLEIQ	168	
Db	88	GIGENLGAELVVEGCVKRR-----	FNTSLIDKSGKIGYKRYKTHL-----	PQH	1322	
QY	159	-----GMR-----	EENNYKRGSTLERPVKKT	TRAGKGLACIDYMRRELS	KLAAQAGAILT	220
Db	133	KEYEAYRPRQHLERKYFERBGLGFRVYUDVADKAKGMPICIDRRMR	REARMVYGLSGAIIIC		1922	
QY	221	-----YPSAGSVGPAHMEVL-----	-DRAAIESCCVLIAAOSG	RIHNETRASHGSMV	2711	
Db	193	GGYNTPTPHNPTRQHDHILTSFNHLSMQAGSYQNGAMSAAAKGVM	-EENOMILGHSCIV		251	
QY	272	DPMGTVARCSG--GRGLSLAIDHNLQOMKONLRFV--QNR	RRPLUG		316	
Db	252	APTEIYALTTTLEDEYTAUVDLDRCKREKHNLEFNKROHOPONG			298	

RESULT 2
US-08-294-871A-64
; Sequence 64, Application US/08294871A

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiko

APPLICANT: Namba, Hirokazu

APPLICANT: Takano, Masayuki

APPLICANT: Yajima, Kazuyoshi

APPLICANT: Yamada, Yukio

APPLICANT: Takahashi, Satomichi

APPLICANT: Okubo, Kazuma

```

APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiratah, Yoshio
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentm Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4008448/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4079222/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```

```

Query Match 11.8%; Score 204; DB 2; Length 303;
Quest Local Similarity 25.7%; Pred. No. 4,4e-14;
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12.

QY 65 ELVQOARIGACIACIAPPE-----APDFIARNPAETLLLESPINDGLLGOYSQIAREC 116
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 29 DMLTGAASRGANFIYEPPELALTFPPRMHFTDEAELDSYTEHE-MGPVRYRLFEKAAL 87
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 117 GIWLSL-----GGFHERGODHONOKIYCHVLLNSKSGSVASYSKTHLCVDEIPGO 168
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 88 GIGFNIIGVAELVEEGVKRR-----FNTSILVKSCKIYVKYRIHL-----PGH 132
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 169 -----GPMR--ESNVTKTCGTLFPPVKTTPACKVGIAICIDMRPPELSLKLQAQAEILT 220
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 133 KEYEYIRPPHLEKRIYFEEGDLGFVVYDVDAKKMGKFIQNDRRMPAMRVMGIRGAELIC 192
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 221 YPSAFGSVYGPAPMVEVL-----LRARAIESQCYVIAAOCGRHHTPRASYG 266
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 193 -----GGVNTPTPHNDIIVPQHDLTSPHLLSMQASGYONGASAAAGKGM--EENOMLLG 246
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 267 HSMVYDPGTVVVARCSE-GRGICLARIDLIHFIQOMROHLPV-QHRRPPLG 316
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 247 HSCIAVAPGEIYALTTLEDEVITAYAVDLDRCDREHIFENFKOHOPOHYG 298
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
US-08-876-398A-26
; Sequence 26, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masaayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBANYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876.398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION/DOCKET NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: (202) 672-5300
: TELEFAX: (202) 672-5399
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 303 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-876-398A-26

Query Match      11.8%; Score 204; DB 3; Length 303;
Best Local Similarity 26.5%; Pred No. 8.4e-14;
Matches 76; Conservative 46; Mismatches 113; Indels 52; Gaps 12.

QY 65 ELVDEARLGLACLAFLPE-----AFDFIARNPAETLLSEPLNGDLGYSQIARBC 116
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 29 DMLTFAASGAFNYIPEPELATLTPPRMHTDEALDSYTEHE-MGPRVVRPLFEKAAL 87
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 117 GIWLSL-----GCFHERGODWEONOKIYCHVLNKSQSVASRYRTHLCDVEIPQO 168
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 88 GIGFNLGVAELVEEGVKRR-----FMNLSIYDKSGKIYGVKRIHL-----PCH 132
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 169 -----GPMR--ESNVTKRGGLPEPPVYKTPACKVGLATIGDYMREPRLSLKLAQAGAEILT 220
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 133 KEYEYIRRPQHLKRTFEFGDLGFPYVDYDAKMGAFICNDKRWPEAMVYGLRGAEITC 192
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 221 --YPSAFSGSVTPGPAWMEVY-----LRARATIESQCYVIAAOCGRHNHETRASYSCHSVV 271
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 193 GGYNTPHTNHPYTPQDHLHSFHHLSMQGSGYQNGAMWSAAKVGW-EENCMLHLSGICIV 251
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 272 DPMGVTVARCSG-GEGLCLARIDHFLQOMRQHLPLYF-QHRPRDLYG 316
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 252 APTGEIVALTTLLEDEVITAAVDLDRCRELREHIFNFKOROPHYG 298
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 4
US-08-876-398A-64
: Sequence 64, Application US/08876398A
: Patent No. 6083752
: GENERAL INFORMATION:
:   APPLICANT: IKENAKA, Yasuhiro
:   APPLICANT: NAKABA, Hirokazu
:   APPLICANT: TAKANO, Masayuki
:   APPLICANT: YAJIMA, Kazuyoshi
:   APPLICANT: YAMADA, Yukio
:   APPLICANT: TAKAHASHI, Satomi
:   TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
:   NUMBER OF SEQUENCES: 70
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: FOLEY & LARDNER
:     STREET: 3000 K Street, N.W.
:     CITY: Washington
:     STATE: D.C.
:     COUNTRY: U.S.A.
:     ZIP: 20007-5109
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/876,398A
:     FILING DATE: 16-JUN-1997
:   CLASSIFICATION: 435
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: US 08/211,641
:     FILING DATE: 11-APR-1994
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: WO PCT/JP93/01101
:       FILING DATE: 05-AUG-1993
:     PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25, 258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-876-398A-64

Query Match 11.8%; Score 204; DB 3; Length 303;
Best Local Similarity 25.7%; Pred. No. 8.4e-14;
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

OY 65 ELVOEARLGAFLPE-----APDFIARNPAETLLSEPLNGDLGGYSQIAREC 116
DB 29 DMLTFAASRGANFIPELALTFPPRMHFTDEALDSYTE-MSGPVVRPLFEKAEL 87
OY 117 GIWLSL-----GGFERGODWONOKIYNCHVLLNSKGSVVASYRTHLCDVEIPQ 168
DB 88 GIGFNLGYALVELVEGVKKR-----FNTSLIVDKSKGIYVKYRIHL-----PGH 132
OY 169 -----GPMR--ESNNTKRGGLTPPVKTPACKVGLATCYDMRPFELSLKLAQAGAEILT 220
DB 133 KTEYATRPQHLKRFEEGDLGFPYVDVDAKMGAFICNDKRWPEAMRWGLRGAELITC 192
OY 221 YPSAGSVYGAHMEVL-----LRARIESQCYVIAAOCGRHETFRASYG 266
DB 193 -----GGYNTPTNPPIVPHDHLTSPFHLLSMOAGSYNGAMSAAGKGM--EENMMLG 246
OY 267 HSMYVDPKCTVVARCSE-GPGICLARIDHFLQOMRQHLPVF-QHRRLPLYG 316
DB 247 HSCIVAPGEIYALTTLEDEVITAAVDLRCRELRHIFNFKHROPOHYG 298

RESULT 5

US-08-294-871A-40
Sequence 40, Application US/08294871A
Patent No. 5824522

GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nambu, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraishi, Yoshio
TITLE OF INVENTION: Immobilized Enzyme Preparation and
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294, 871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25, 258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-40

Query Match 11.7%; Score 203; DB 2; Length 303;
Best Local Similarity 25.7%; Pred. No. 1.1e-13;
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

OY 65 ELVOEARLGAFLPE-----APDFIARNPAETLLSEPLNGDLGGYSQIAREC 116
DB 29 DMLTFAASRGANFIPELALTFPPRMHFTDEALDSYTE-MSGPVVRPLFEKAEL 87
OY 117 GIWLSL-----GGFERGODWONOKIYNCHVLLNSKGSVVASYRTHLCDVEIPQ 168
DB 88 GIGFNLGYALVELVEGVKKR-----FNTSLIVDKSKGIYVKYRIHL-----PGH 132
OY 169 -----GPMR--ESNNTKRGGLTPPVKTPACKVGLATCYDMRPFELSLKLAQAGAEILT 220

Db 133 KEYEAYRPFQHLERKRYEPDGLGFPYVDAAKMGFICNDRRPEAMRVMLGAEIIC 192
QY 221 YPSAFSGVTGPAHMEVL-----LRAAIESQCYVIAAOCGRHHETRASYG 266
Db 193 -----GGYNTPTNPVLPOHDHLTSFHLLSMOAGSYONGAMSAAGKAGM-BEENCMLLG 246
QY 267 HSNVDPWGTIVARCS-GRGLCIARIDLHFLQOMROHLPVF-OHRRPDLYG 316
Db 247 HSCIVAPTGEIVALTTLLEDEVITAAYDLDRCHRELREHIFNFKOHPQHYG 298

RESULT 6
US-08-876-398A-40
; Sequence 40, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKABA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,758
; FILING DATE: 12-APR-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEO ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-876-398A-40

Query Match 11.7%; Score 203; DB 3; Length 303;
Best Local Similarity 25.7%; Pred. No. 1,1e-13;
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

QY 65 ELVQEARLACLAFLPE-----AFDEIARNPAETLLSEPLNDGLGOYSQALREC 116
Db 29 DMLTKAASRGANFLVPELALITTFPRMHTDELDSEFETE-MPGPVYRPLFEKAAEL 87
QY 117 GIWLSL-----GCFHERGQDWEQONKIYNCHVILNSKGSVYASRKTHLDVLEIPQ 168
Db 88 GIGFNLCYAEILVVEGVKRR-----FNSILVDKSGKIKYGRKTHL-----PGH 132
QY 169 -----GPMR--ESNTRKPGSTLEPPYKTPAGKVGLAICTDMRPELSKLDAQGAELIT 220
Db 133 KEYEAYRPFQHLERKRYEPDGLGFPYVDAAKMGFICNDRRPEAMRVMLGAEIIC 192
QY 221 YPSAFSGVTGPAHMEVL-----LRAAIESQCYVIAAOCGRHHETRASYG 266
Db 193 -----GGYNTPTNPVLPOHDHLTSFHLLSMOAGSYONGAMSAAGKAGM-BEENCMLLG 246
QY 267 HSNVDPWGTIVARCS-GRGLCIARIDLHFLQOMROHLPVF-OHRRPDLYG 316
Db 247 HSCIVAPTGEIVALTTLLEDEVITAAYDLDRCHRELREHIFNFKOHPQHYG 298

RESULT 7
US-08-294-871A-6
; Sequence 6, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKABA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; APPLICANT: OKUDO, Kazuma
; APPLICANT: YAMADA, Kazuhiko
; APPLICANT: HIRASHI, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,758
; FILING DATE: 12-APR-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEO ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-294-871A-6

Query Match 11.7%; Score 203; DB 3; Length 303;
Best Local Similarity 25.7%; Pred. No. 1,1e-13;
Matches 75; Conservative 45; Mismatches 110; Indels 62; Gaps 12;

```

APPLICATION NUMBER: JP 078840/1991
  FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 140051/1991
    FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: PCT/JP92/00739
    FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 212692/1992
    FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: PCT/JP93/01101
    FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 340078/1992
    FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
  NAME: Wegner, Harold C.
  REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (202) 672-5300
  TELEFAX: (202) 672-5399
  TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
  LENGTH: 303 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
MOLECULE TYPE: protein
OS-08-294-871A-6

```

Query Match	11.6%;	Score 202;	DB 2;	Length 303;
Best Local Similarity	25.7%;	Pred. No. 1.4e-13;		
Matches	75;	Conservative	45;	Mismatches 110;
				Indels 62;
				Gaps 12

```

OY 65 ELVOEAAITGACIAFLPE-----AFDFIARNAPEITLLSEPLNDDIGQYSOLAREC 116
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 29 DMLTKRASRGANFIYFPELALITTFPRKMHFTDEALDELDSEFETE-MGPPVVRPLFEKAEL 87

OY 117 GIWLSL-----GGEHENGQDWEONOKIYNCHVLLNSKGSVVASYRKTHLCDVEIPGO 168
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 88 GIGENLGAELVGEVGWYR-----FNTSILVDKSGKIYKGKRIIHL-----PGH 132

OY 169 -----GPMR--EENHYRKGGTLEPRPKYTRAGKGLAIQYDMRPELSLTLAAGAFILR 220
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 133 KEYEAYRPFQHLKRYFERPEGLGFPPYDVADAKKGMFTCDRRKREPMRWGLRGATIIIC 192

OY 221 YPSAFGSVYGAHNEVL-----LRAATIESOCYVIAAQAOCGRNHETRASGY 266
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 193 -----GGYNFTNPIRVDQHDHLSFPHLLLSMQGSGYNGAMSAAKGVGM-EENCMLLG 246

OY 267 HSMVYDPRGIVYARCSE-GPGLCLARIDLHFLQDMQHRLPVF-OHRPRDLYG 316
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 247 HSCIVAPRGELVALTTLEDEVITAAVLDRCRELRHINFNOKRORPOHYG 298

```

RESULT 8
 US-08-876-398A-6
 : Sequence 6, Application US/08876398A
 : Patent No. 6083752
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: IKENAKA, Yasuhiko
 : APPLICANT: NANBA, Hirokazu
 : APPLICANT: TAKANO, Masayuki
 : APPLICANT: YAJIMA, Kazuyoshi
 : APPLICANT: YAMADA, Yukio
 : APPLICANT: TAKAHASHI, Setomi
 :
 : TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
 : TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
 :
 : NUMBER OF SEQUENCES: 70

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 216992/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-876-398A-6

```

```

Query Match          11.6%; Score 202; DB 3; Length 303;
Best Local Similarity    25.7%; Pred.No. 1.4e-13;
Matches      75; Conservative   45; Mismatches 110; Indels     62; Gaps     12.

QY       65 ELVDAAALGACLAFLPE-----ADFIAKNAEETLLSEPLNDLIGYSQLAREC 116
           ::::| | | | | | | | | | | | | | | | :
DB       29 DMLTAAASRGANFIVEPRLATTFPRRHFTDEAELDSFEYE-MGPVAVRLFEEKAAL 87
           ::::| | | | | | | | | | | | | | | | :
QY       117 GIWLSL-----GGFHREGDWEQNOKIYNCHVLNLSKGSVVASRYRTHLCDAVEIQ 168
           || | | | | | | | | | | | | | | | | |
DB       88 GIGFWLGIAELVEVGGVKR-----FHTSLIVDSGKIYGRKHIL-----PGH 132
           || | | | | | | | | | | | | | | | | |
QY       169 -----GPMR--ESNYTPRGGLTEBPVKTRPAKVGLAIICYDRFPSELISKLQAQAEILT 220
           ::::| | | | | | | | | | | | | | | | :
DB       133 KEYEAYRFQHLEKRYFFEGDLGFPPYUDADAKMGAFICNDNRWMEAMVMVLRGAELIT 192
           ::::| | | | | | | | | | | | | | | | :
QY       221 YPSAGSVTGPAHWLEV-----LRARAIESOCYVIANAACSRRHHETRASYG 266
           ::::| | | | | | | | | | | | | | | | :
DB       193 -----GGINTPTHNPLVPQHHDLTFSFNHLISMQASSYONGAASAAGKKGM-EENCMI LG 246
           ::::| | | | | | | | | | | | | | | | :
QY       267 HSMVDDPMGYTVARCS-GPGCLARIIDLHPLOQRONHPV-QHRRPDYLG 316
           || | | | | | | | | | | | | | | | | |
DB       247 HSCIAPAGEIVAALTTLDEEVITTAAYVDLRCRELREHIENFKORHORONYG 298
           || | | | | | | | | | | | | | | | | |

```

```

GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuniko
APPLICANT: Hiraishi, Yoshiko
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136

```

```

; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-294-871A-44

Query Match 11.6%; Score 201.5; DB 2; Length 303;
Best Local Similarity 25.8%; Pred. No. 1,6e-13;
Matches 77; Conservative 44; Mismatches 115; Indels 63; Gaps 13;

QY 50 VTSFKNQENFKTCALVQZARLGAQLAPLEAFDIARPAFTLLSPFLNDLLGQY 109
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 LTKASRGANFVPEPL-----ALTFFPRVY-FTDEAEIDSFYETE-MPGVVRPL 80
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 SOLARECGIMVLSL-----GGFHERGQDWEQNOVKLYNCVLLNSKGSVASYRKTSLC 161
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 FPKAEIGIGFNLGYALVYEGYKRR-----FNSILYDKSGKIVGKRTKTHL- 129
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 DVEIPGQ-----GPMR-ESNYTKPGTLEPPVKTGAGVGAICYDMRPPLSLKLAQ 213
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 ---PGHKEYEAYRPFQHLERKRYEPDGLGPPYVDVDAKMGMTICNDRMRPEAMRVML 185
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 214 AGAEILTPSAFGSVTGPANWEVL-----LRAAIESOCYVIAAQCGRHH 259
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 RGAELIC-----GGYNPTNPPLVPOHDHLTSFHHLLSMQAGSYQNGAWSAAAGKGM-E 239
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 ETRASGSHMVDPMGTIVARCSE-GPGLCLARIDLHFLQOMROHLPVF-OHRRPDLYG 316
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 ENCMLGHSCTVAPTGETVALTTLEDEVTITAAVDLDRCRRLRSHITFNKORHPQHYG 298
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-08-876-398A-44
; Sequence 44, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAMBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992

```


APPLICANT: Nanba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraishi, Yoshio
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294, 871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971, 758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917, 111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211, 641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:

LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-34

Query Match 11.5%; Score 200.5; DB 2; Length 303;
Best Local Similarity 25.8%; Pred. No. 2e-13;
Matches 77; Conservative 44; Mismatches 115; Indels 63; Gaps 13;

QY 50 VTSPNKKQENFKCAELVQBAARLGAFLPEAFDFIARNPATLLISPLNGDLGQY 109
DB 31 LTRASRGANFYVPEL-----ALTTFPRMY-FTDEHEDSFETE-MQPVVRL 80
QY 110 SQLARECGIWLST-----GCFHERGQWQONQKIYNCHVILNKGSVVASYRKTLC 161
DB 81 FEKAHELIGIFNLGYALVEGVKRR-----FNTSILVDKSGKIYGYRKTIL- 129
QY 162 DVEIPGQ-----GPMR--ESNITKPGTLEPPYKTAGVGLAICYDMRPELSLKAQ 213
DB 130 ---PGHKEYEAYRPFQHLKRYFEPDGLGPVYVDAAKMGFICNDRWPEAMRVGL 185
QY 214 AGAEILTFPSAFSGSVTGAHMEVL-----LRRALESQCYVIAAOCGRNH 259
DB 186 RGAELIC-----GCYNTPTHNPVLVPOHDHLSFMHLSMAGSIQNGAMSAAGKVG-M-E 239
QY 260 ETRASVGHSMVVDPMGTIVVARS-GRGLCLARIDLHFLQMRQHLVF-QHRRPDLVG 316
DB 240 ENCMLLGHSCIVAPGTGEIVALTTLDEDEVITAAVDLDRCRELRFHIFNQHRQPHYG 298

Search completed: April 26, 2002, 16:55:12
Job time: 255 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:55:12 ; Search time 28.88 Seconds
(without alignments)
282.070 Million cell updates/sec

Title: US-09-357-675c-25

Perfect score: 1921
Sequence: 1 PLAAXLADRPDPRLRMV.....LPVGHRRPDLVGNLGHPLS 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 212252 segs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCrUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.5	10.4	303	2	US-08-294-871A-64
2	200.5	10.4	303	3	US-08-876-398A-64
3	199.5	10.4	303	2	US-08-294-871A-26
4	199.5	10.4	303	2	US-08-294-871A-40
5	199.5	10.4	303	2	US-08-294-871A-44
6	199.5	10.4	303	2	US-08-876-398A-26
7	199.5	10.4	303	3	US-08-876-398A-40
8	199.5	10.4	303	3	US-08-876-398A-44
9	198.5	10.3	303	2	US-08-294-871A-6
10	198.5	10.3	303	2	US-08-294-871A-34
11	198.5	10.3	303	2	US-08-294-871A-60
12	198.5	10.3	303	3	US-08-876-398A-6
13	198.5	10.3	303	3	US-08-876-398A-34
14	198.5	10.3	303	3	US-08-876-398A-60
15	197.5	10.3	303	2	US-08-294-871A-62
16	197.5	10.3	303	3	US-08-876-398A-62
17	196.5	10.2	303	2	US-08-294-871A-10
18	196.5	10.2	303	2	US-08-294-871A-12
19	196.5	10.2	303	2	US-08-294-871A-22
20	196.5	10.2	303	2	US-08-294-871A-38
21	196.5	10.2	303	2	US-08-294-871A-42
22	196.5	10.2	303	2	US-08-294-871A-46
23	196.5	10.2	303	2	US-08-294-871A-54
24	196.5	10.2	303	2	US-08-294-871A-58
25	196.5	10.2	303	2	US-08-294-871A-68
26	196.5	10.2	303	3	US-08-876-398A-10
27	196.5	10.2	303	3	US-08-876-398A-12

28	196.5	10.2	303	3	US-08-876-398A-22	Sequence 22, Appl
29	196.5	10.2	303	3	US-08-876-398A-38	Sequence 38, Appl
30	196.5	10.2	303	3	US-08-876-398A-42	Sequence 42, Appl
31	196.5	10.2	303	3	US-08-876-398A-46	Sequence 46, Appl
32	196.5	10.2	303	3	US-08-876-398A-54	Sequence 54, Appl
33	196.5	10.2	303	3	US-08-876-398A-58	Sequence 58, Appl
34	196.5	10.2	303	3	US-08-876-398A-68	Sequence 68, Appl
35	195.5	10.2	303	1	US-07-917-111-2	Sequence 2, Appl
36	195.5	10.2	303	1	US-07-917-111-3	Sequence 3, Appl
37	195.5	10.2	303	1	US-08-479-638-2	Sequence 2, Appl
38	195.5	10.2	303	1	US-08-479-638-3	Sequence 3, Appl
39	195.5	10.2	303	2	US-08-294-871A-2	Sequence 2, Appl
40	195.5	10.2	303	2	US-08-294-871A-4	Sequence 4, Appl
41	195.5	10.2	303	2	US-08-294-871A-8	Sequence 8, Appl
42	195.5	10.2	303	2	US-08-294-871A-14	Sequence 14, Appl
43	195.5	10.2	303	2	US-08-294-871A-16	Sequence 16, Appl
44	195.5	10.2	303	2	US-08-294-871A-18	Sequence 18, Appl
45	195.5	10.2	303	2	US-08-294-871A-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-294-871A-64
Sequence 64, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-APR-1994
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/921,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990

;; FILING DATE: 27-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 078840/1991
;; FILING DATE: 11-APR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 140051/1991
;; FILING DATE: 12-JUN-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP92/00739
;; FILING DATE: 10-JUN-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 212692/1992
;; FILING DATE: 10-AUG-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/JP93/01101
;; FILING DATE: 05-AUG-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 340078/1992
;; FILING DATE: 21-DEC-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wegner, Harold C.
;; REGISTRATION NUMBER: 25,258
;; REFERENCE/DOCKET NUMBER: 74129/137/AOPA
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 672-5300
;; TELEFAX: (202) 672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 64:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 303 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-294-871A-64

Query Match 10.4%; Score 200.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred. No. 1.6e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;
QY 83 LVAVCQ--VTSTPDKQONFTCAELVREARLGACLAFLDE--AFDFIAR---DPAET 133
DB 5 ILAVGQGPRIARAEIRREVVVRLDMLTKASRGANFTVPELALTTFFPRNHFDDEL 64
QY 134 LHLSE-PLGSKLLEBYTOLARECGIMLSLGFH---ERGDMEQOTKYINCHVLLNSGA 189
DB 65 DSFETEMPGVVRPLFEKAEALGIGFNLGYAELVVEGG---VKRRFNTSILVDKSGK 119
QY 190 VVATYRKTHLCDVEIRPGGPRCESNSTMPGSLSPVSTP-----AGKIGLAVC 238
DB 120 IVGKYRKITHL-----PGHK---EYEAYRPFQHLERKRYEPDGLGPRVYDVDAKKMGFTIC 171
QY 239 YDMRPPELSIALAOGAAILTPSAFGSITGPANHEVL-----LRAALET 284
DB 172 NDRMRPEAMRWGRLGAEIIC-----GGYNTPTNPIYRPQDHILTSFHHLLSMOAGSYON 226
QY 285 QCYVVAALAAOCGRHHEKRAVSYGSWVDPGWIVVARGSE--GGLCLARIIDLNYLRLRHH 343
DB 227 GAWSAALAGKVGW--EENCMLGSHCSIVAPTGEIVALTTLLEDEVITAAVDDLRCRELREHI 285
QY 344 PVF-QHRRPDLVG 355
DB 286 FNEKQHRQPHYHG 298

RESULT 2
US-08-876-398A-64
; Sequence 64, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAMBA, Hirokazu
; APPLICANT: TANAKA, Masayuki

;; APPLICANT: YAJIMA, Kazuyoshi
;; APPLICANT: YAMADA, Yukio
;; APPLICANT: TAKAHASHI, Satomi
;; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
;; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
;; NUMBER OF SEQUENCES: 70
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FOLEY & LARDNER
;; STREET: 3000 K Street, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/876,398A
;; FILING DATE: 16-JUN-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/211,641
;; FILING DATE: 11-APR-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/JP93/01101
;; FILING DATE: 05-AUG-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 340078/1992
;; FILING DATE: 21-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 212692/1992
;; FILING DATE: 10-AUG-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wegner, Harold C.
;; REGISTRATION NUMBER: 25,258
;; REFERENCE/DOCKET NUMBER: 74129/130
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 672-5300
;; TELEFAX: (202) 672-5399
;; INFORMATION FOR SEQ ID NO: 64:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 303 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-876-398A-64

Query Match 10.4%; Score 200.5; DB 3; Length 303;
Best Local Similarity 26.2%; Pred. No. 1.6e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;
QY 83 LVAVCQ--VTSTPDKQONFTCAELVREARLGACLAFLDE--AFDFIAR---DPAET 133
DB 5 ILAVGQGPRIARAEIRREVVVRLDMLTKASRGANFTVPELALTTFFPRNHFDDEL 64
QY 134 LHLSE-PLGSKLLEBYTOLARECGIMLSLGFH---ERGDMEQOTKYINCHVLLNSGA 189
DB 65 DSFETEMPGVVRPLFEKAEALGIGFNLGYAELVVEGG---VKRRFNTSILVDKSGK 119
QY 190 VVATYRKTHLCDVEIRPGGPRCESNSTMPGSLSPVSTP-----AGKIGLAVC 238
DB 120 IVGKYRKITHL-----PGHK---EYEAYRPFQHLERKRYEPDGLGPRVYDVDAKKMGFTIC 171
QY 239 YDMRPPELSIALAOGAAILTPSAFGSITGPANHEVL-----LRAALET 284
DB 172 NDRMRPEAMRWGRLGAEIIC-----GGYNTPTNPIYRPQDHILTSFHHLLSMOAGSYON 226
QY 285 QCYVVAALAAOCGRHHEKRAVSYGSWVDPGWIVVARGSE--GGLCLARIIDLNYLRLRHH 343
DB 227 GAWSAALAGKVGW--EENCMLGSHCSIVAPTGEIVALTTLLEDEVITAAVDDLRCRELREHI 285

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/294,871A
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-40

Query Match 10.4%; Score 199.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred. No. 2.1e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY 83 LVAVCQ---VSTPDKQONFKTCAELVREARLGACLAFLPE--AFDTAR---DPAET 133
Db 5 ILAVQCGPPIARAETRECVVLLDMLTKRASRGANFIVFPELATVTFPRMHTDEAEL 64

QY 134 LHLSP-PLGGKLLREYTLQARCGGLMSLGGPH---EKGQDWEQYOKIYNCHVLLNSKA 189
Db 65 DSYETEMPGPVPRPLFEKAELGIGFNLGYAELVEGG---VKRRNTSILVDSGK 119

QY 190 VVATYRKTHLCDVEIPGCGPCESNSTMPGSLSPVSTP-----AGKIGLAVC 238
Db 120 YGKTKRKHL-----PGHK---EYEA YRPQHLERKRYEPGDLGFPVYDVDAKMGMFIC 171

QY 239 YMRPELSLALAQAGALITPSPAFGSITGPANHVEL-----LRARAIT 284
Db 172 NDRMRPEAMRWVGLRGALITC-----GGYNTPTHNPVQHDHLSFHHLSMQAGSYQN 226

QY 285 OCYVVAACGGRHNEKRASYGSMVVDPMGTAVARCSE-GPGLCLARIDLNYLRQLRHL 343
Db 227 GAWSAAGKAGH-EENCMLGHSCIVAPGELVALTTLEDEVITAAVDLDRCRELREHI 285

QY 344 PVF-QHRPDLVG 355
Db 286 FNEKHQHPQHYG 298

RESULT 5
US-08-294-871A-44
Sequence 44, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Hirataishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/294,871A
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991


```

      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/J991/01696
      FILING DATE: 06-DEC-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 407922/1990
      FILING DATE: 27-DEC-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 078840/1991
      FILING DATE: 11-APR-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 140051/1991
      FILING DATE: 12-JUN-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/J992/00739
      FILING DATE: 10-JUN-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 212692/1992
      FILING DATE: 10-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/J993/01101
      FILING DATE: 05-AUG-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 340078/1992
      FILING DATE: 21-DEC-1992
      ATTORNEY/AGENT INFORMATION:
      NAME: Wegner, Harold C.
      REGISTRATION NUMBER: 25,258
      REFERENCE/DOCKET NUMBER: 74129/127/AOPA
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202) 672-5300
      TELEFAX: (202) 672-5399
      TELEX: 904136
      INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 303 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      US-08-294-871A-6

Query Match          10.3%; Score 198.5; DB 2; Length 303;
Best Local Similarity 26.2%, Pred. No. 2.7e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;

QY   83 LVAVO---VNSTPKDQONFKTCALVEAEARTLGAFLPE--AFDFIAR----DPDET 133
    ::::| | | | : : | | : : : | | | | | | | | | | | | | | | | |
DB   5 ILVGQGCGPIRAETREOVVVRLDLMLTKASRGANFIYFELALITFFPRRHFTDEALE 64
    ::::| | | | : : | | : : : | | | | | | | | | | | | | | | | |

QY   134 LHLSE-PGSGKLLLEYTDLARECGILMLSGFH---ERGQDEWOTOKLYNCVALLNSKA 189
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   65 DSYFTENPGRPVRRFLFEKAELGI GFNLGYALVVEGS-----VKRRNTSILVDKSK 119
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   190 VVAITYRKTHLCDVELIPGGCPMCESNSTMPSGESPSVTSP-----AGKTGLAVC 238
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   120 IVGKKRIKTHL-----PGHK---EYEAYRPFOHLERKYEPFGDIGFPVYDVDAKKGMFIC 171
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   239 YDMRPRELIALQAQCAELLTPSAFGSTGTGAHNEVL-----LRARIET 284
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   172 NDRWRPEAMRWVGLRGAEITC-----GGYNTPTNHPLVQHDLHSFNHLLSMQSGSYQN 226
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   285 QCYUVAALAQCGRHNHRKASYGSHSVVDWGVVAACSE-GPGICLARIDLTNYLRRLH 343
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   227 GAMSAAAKVGC-EENCMLHGSHCI VAPTGEIVALTTLTEDEVITAAYDLDCRFLEHNI 285
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   344 PVF-OHRRPDLYG 355
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   286 FNFKOHROPNIYG 298
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
US-08-294-871A-34
Sequence 34, Application US/08294871A

```

Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraishi, Yoshiko
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399

TELEX: 904136
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-34
Query Match 10.3%; Score 198.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred. No. 2.7e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;
QY 83 LVAVCO---VTPDROKQNEKTCALVREARGLACLAFLPE--APDFLAR---DPAET 133
DB 5 ILAVGCGPIARARETRQVAVRLLDMLTAAASGANFIYFPEALTLTFPPRWTFDEABL 64
QY 134 LHLSE-PLGKGLLEEYQTALRECGMLSLGGEH---ERGDMEQITOKIYNCHVLLNSKA 189
DB 65 DSFEYTEMFGPVVRRPLFEKAAELGIGNLGYALVEGG---VKRRNTSILVDKSGK 119
QY 190 VVATYRKTLCVETIPGCGPMCESNTMPGSLSEVSTP-----AGKIGLAVC 238
DB 120 IVGKYRKRIH---PGHK---EYEAIRPFQHLKRYFEPGDLGFPVYVDADAKMGMFTC 171
QY 239 YDMRPEPLSLALAQAGAEILTPSAGFSITGPAHMEVL-----LRARATET 284
DB 172 NDRMPPEARVMKLRGAELIIC-----GYNTPTNPNLVQHDLTFSFHLLSMQAOSYON 226
QY 285 QCVYVAAQCGRRHERRASYGSMVVDPMGTVAARCSE-GPGICLARIDLNTLRQLRRHL 343
DB 227 GAWSAAGKVGW-EENCMLGHSCIVAPGEIYALTTLEDEVITFAAVDLDCREIREHI 285
QY 344 PVF-QHRPDLTG 355
DB 286 FNEKHROPQHYG 298
RESULT 11
US-08-294-871A-60
Sequence 60, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraishi, Yoshiko
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-294-871A-60

Query Match 10.3%; Score 198.5; DB 2; Length 303;
Best Local Similarity 26.2%; Pred. No. 2,7e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;
QY 83 LVAVCQ---VTSTPDKQONFTCAELVREARLACLAFLPE--AFDFIAR---DPAET 133
DB 5 ILAVGQGGPIARAETREQVAVRLLDMLTKAASRGANFIVPELALTTFFPRMHTDEAEL 64
QY 134 LHUSE-PLGSKLLEBYTQLAREGLMLSLGFGH--ERGDWEQOTOKLYNCHVLNLSGA 189
DB 65 DSFEYTEMPPGVVRLPEKKAELGIGRLGALVELVEGG-----VKRFNLSILVDKSGK 119
QY 190 VVATYRTHLCDVEIPGQGPWCESNTPGSPLESVPSTP-----AGKIGLAVC 238
DB 120 IVGKYRKILH-----PGHK---EYEAAYRPFQHLKRYFEPDGLGFPYVDVDAKMGFIC 171
QY 239 YDMFPEPLSLALAGAGAILTYPSAFGSITPRAMWEVL-----LARALET 284
DB 172 NDRMWPAMRWYMGRLGAEILIC-----GGYNTPTNHPPLVPQDHLTFSHLLSMDAGSYON 226
QY 285 QCYVVAAGCGRHHEKBSYSGHSMVVDPMGTIVVARCSE-GRGLCLARIDLNYLQLRRLH 343

DB 227 GAWSAAGKSGM-BENCMLLGHSCIVAPTGEIVALTTLLEDEVTAAVDLDRCRELRHI 285
QY 344 PVF-QHRRPDLXG 355
DB 286 FNFKOHROPQHYG 298

RESULT 12
US-08-876-398A-6
Sequence 6, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NAKABA, Hirokazu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-876-398A-6

Query Match 10.3%; Score 198.5; DB 3; Length 303;
Best Local Similarity 26.2%; Pred. No. 2,7e-13;
Matches 82; Conservative 46; Mismatches 126; Indels 59; Gaps 14;
QY 83 LVAVCQ---VTSTPDKQONFTCAELVREARLACLAFLPE--AFDFIAR---DPAET 133
DB 5 ILAVGQGGPIARAETREQVAVRLLDMLTKAASRGANFIVPELALTTFFPRMHTDEAEL 64

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:23 ; Search time 57.61 Seconds
(without alignments)
919.121 Million cell updates/sec

Title: US-09-357-675C-25

Perfect score: 1921

Sequence: 1 PLAAAXLAPDRPRDRLRMV.....LPVFQHRPRDLGNLGHPLS 362

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1743	90.7	327	4	076091
2	1482	77.1	323	11	088526
3	1479	77.0	323	11	09114
4	1372	71.4	290	11	09DBF2
5	956	49.8	288	13	0918G0
6	710.5	37.0	460	5	076464
7	702	36.5	440	5	076463
8	645.5	33.6	316	10	091E50
9	575.5	30.0	276	3	094660
10	537.5	28.0	272	1	027839
11	497	25.9	282	2	09HVU6
12	485.5	25.3	276	4	09NORA
13	479	24.9	270	2	083040
14	474.5	24.7	276	11	09JHW2
15	436.5	22.7	275	2	09KU04
16	433.5	22.6	283	2	09A9Y3
17	406	21.1	318	10	09LYH1
18	401	20.9	270	2	09J713
19	398	20.7	270	2	09K0W5

20	397	20.7	113	6	029367	029367 sus scrofa
21	395.5	20.6	189	11	09CTG9	09ctg9 mus musculus
22	387	20.1	283	5	09VHE4	09vhe4 drosophila
23	371	19.3	231	11	09DOS1	09dos1 mus musculus
24	342.5	17.8	264	2	09XA70	09xa70 streptomyces
25	318	16.6	297	2	09ARE7	09are7 delinococcus
26	317	16.5	270	1	09IWM8	09iwm8 thermoplasma
27	312	16.2	257	1	030121	030121 archaeoglob
28	291.5	15.2	300	10	09XG19	09xg19 lycopersico
29	279	14.5	287	1	09Y9L1	09y9l1 aeropyrum p
30	274.5	14.3	271	2	09KE11	09ke11 bacillus ha
31	273.5	14.2	292	2	0916J8	0916j8 pseudomonas
32	264.5	13.8	295	2	09PA02	09pa02 xyella fas
33	262	13.6	270	2	091241	091241 pseudomonas
34	255.5	13.3	259	2	031664	031664 bacillus su
35	253.5	13.2	282	2	09Z505	09z505 zymomonas m
36	252	13.1	298	12	089413	089413 paramedium
37	249.5	13.0	280	2	09ABU5	09ab15 caulobacter
38	248	12.9	292	2	09ABU5	09ab15 caulobacter
39	246.5	12.8	280	2	069939	069939 streptomyces
40	241	12.5	220	10	09ZQH4	09zqh4 arabidopsis
41	240	12.5	220	2	025452	025452 helicobacter
42	239	12.4	292	2	09ZL86	09zl86 helicobacter
43	238.5	12.4	262	2	091104	091104 streptomyces
44	233.5	12.2	271	2	09CB47	09cb47 mycobacteri
45	230.5	12.0	294	2	09ZMC7	09zmc7 helicobacter

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	327 AA.
076091	076091			
AC	076091			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	NITRILASE HOMOLOG 1.			
GN	NIT1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98337986; PubMed=9671749;			
RA	Pekarsky Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,			
RA	Tillib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,			
RA	Buchberg A.M., Mazo A., Brenner C., Croce C.M.;			
RT	Nitrilase and Fhl homologs are encoded as fusion proteins in			
RT	Drosophila melanogaster and Caenorhabditis elegans."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).			
DR	EMBL: AF069984; AAC39901.1;			
DR	EMBL: AF069987; AAC39907.1;			
DR	InterPro: IPR001110; UPF0012.			
DR	InterPro: IPR003010; CN_hydrolase.			
DR	Pfam: PF00795; CN_hydrolase; 1.			
DR	PROSITE: PS01227; UPF0012; 1.			
SO	SOURCE 327 AA; 35896 MW; 90F7FB9D4BA627B1 CRC64;			

Query Match 90.7%; Score 1743; DB 4; Length 327;
Best Local Similarity 100.0%; Pred. No. 4.1e-147;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY	37	LGFTTRPHRFLSLICGLRIPQSVLCAPRRPAMAISSSSCCLPLVAVCQVSTPDQK 96
DB	2	LGFTTRPHRFLSLICGLRIPQSVLCAPRRPAMAISSSSCCLPLVAVCQVSTPDQK 61
QY	97	QNFETCAELVREARLACLAFLPEAFDFIARDPAETLHLSPEPLGGRKLEETYLARECG 156
DB	62	QNFETCAELVREARLACLAFLPEAFDFIARDPAETLHLSPEPLGGRKLEETYLARECG 121

```
OY 157 LMLSIGFHERGQDMEQOTKIYNCHVLNSKGAVVATYRKTHLCDVEIPGQPMCESNST 216
      |||||||
Db 122 LMLSIGFHERGQDMEQOTKIYNCHVLNSKGAVVATYRKTHLCDVEIPGQPMCESNST 181

OY 217 MGPSLESFVSTPAGKIGIAGVCDYDMRFPPLSLALQAAGAEIITYPARFSTIGPAHWEVL 276
      |||||||
Db 182 MGPSLESFVSTPAGKIGIAGVCDYDMRFPPLSLALQAAGAEIITYPARFSTIGPAHWEVL 241

OY 277 LRAALETQCYVVAACGCRHHEKRASYGSHSVVDPMGTVVVARCSGPGICLARIIDLNTL 336
      |||||||
Db 242 LRAALETQCYVVAACGCRHHEKRASYGSHSVVDPMGTVVVARCSGPGICLARIIDLNTL 301

OY 337 ROLRRHLVPFOHRRPDLVGNLGHPLS 362
      |||||||
Db 302 ROLRRHLVPFOHRRPDLVGNLGHPLS 327
```

```
RESULT 2
O88526 PRELIMINARY: PRT: 323 AA.
ID O88526;
AC O88526;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE HOMOLOG 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=98337986; PubMed=9671749;
RA Pekarsky Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhl1 homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL: AF069985; AAC40184.1; -.
DR MGD: MGI:1350916; NITL.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA: 35717 MW: 10151CEB151DF2C7 CRC64;
```

```
Query Match 77.1%; Score 1482; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 6.8e-124;
Matches 275; Conservative 27; Mismatches 20; Indels 4; Gaps 2;
```

```
OY 37 LGFTRPPRHRLSLICPGIRIPOLSVLCQOPPRAMAISSSCCELPVAVCVOTSTPDKQ 96
      |||||||
Db 2 LGFTRPPRHQ--LLCTGRLRLTLPVLCQOPPRKMS--SSTSWELPLVAVCVOTSTPDKQ 57

OY 97 QNFKTCALVREARLGAACLAFLPEAFDIARDPAETLILSPGLKLEETTLAREG 156
      |||||||
Db 58 ENFKTCALVREARLGAACLAFLPEAFDIARDPAETLILSPGLKLEETTLAREG 117

OY 157 LMLSIGFHERGQDMEQOTKIYNCHVLNSKGAVVATYRKTHLCDVEIPGQPMCESNST 216
      |||||||
Db 118 LMLSIGFHERGQDMEQOTKIYNCHVLNSKGAVVATYRKTHLCDVEIPGQPMCESNST 177

OY 217 MGPSLESFVSTPAGKIGIAGVCDYDMRFPPLSLALQAAGAEIITYPARFSTIGPAHWEVL 276
      |||||||
Db 178 KPGTLEPVPKTPAGKIGIAGVCDYDMRFPPLSLALQAAGAEIITYPARFSTIGPAHWEVL 237

OY 277 LRAALETQCYVVAACGCRHHEKRASYGSHSVVDPMGTVVVARCSGPGICLARIIDLNTL 336
      |||||||
Db 238 LRAALETQCYVVAACGCRHHEKRASYGSHSVVDPMGTVVVARCSGPGICLARIIDLNTL 297
```

```
OY 337 ROLRRHLVPFOHRRPDLVGNLGHPLS 362
      |||||||
Db 298 ROLRRHLVPFOHRRPDLVGNLGHPLS 323
```

```
RESULT 3
O9R1N4 PRELIMINARY: PRT: 323 AA.
ID O9R1N4;
AC O9R1N4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
GN NITL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=98337986; PubMed=9671749;
RA Pekarsky Y., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhl1 homologs are encoded as fusion proteins in
RT Drosophila melanogaster and Caenorhabditis elegans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR EMBL: AF069988; AAC40185.1; -.
DR MGD: MGI:1350916; NITL.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA: 35705 MW: F8CD7730713665EF CRC64;
```

```
Query Match 77.0%; Score 1479; DB 11; Length 323;
Best Local Similarity 84.4%; Pred. No. 1.3e-123;
Matches 275; Conservative 26; Mismatches 21; Indels 4; Gaps 2;
```

```
OY 37 LGFTRPPRHRLSLICPGIRIPOLSVLCQOPPRAMAISSSCCELPVAVCVOTSTPDKQ 96
      |||||||
Db 2 LGFTRPPRHQ--LLCTGRLRLTLPVLCQOPPRKMS--SSTSWELPLVAVCVOTSTPDKQ 57

OY 97 QNFKTCALVREARLGAACLAFLPEAFDIARDPAETLILSPGLKLEETTLAREG 156
      |||||||
Db 58 ENFKTCALVREARLGAACLAFLPEAFDIARDPAETLILSPGLKLEETTLAREG 117

OY 157 LMLSIGFHERGQDMEQOTKIYNCHVLNSKGAVVATYRKTHLCDVEIPGQPMCESNST 216
      |||||||
Db 118 LMLSIGFHERGQDMEQOTKIYNCHVLNSKGAVVATYRKTHLCDVEIPGQPMCESNST 177

OY 217 MGPSLESFVSTPAGKIGIAGVCDYDMRFPPLSLALQAAGAEIITYPARFSTIGPAHWEVL 276
      |||||||
Db 178 KPGTLEPVPKTPAGKIGIAGVCDYDMRFPPLSLALQAAGAEIITYPARFSTIGPAHWEVL 237

OY 277 LRAALETQCYVVAACGCRHHEKRASYGSHSVVDPMGTVVVARCSGPGICLARIIDLNTL 336
      |||||||
Db 238 LRAALETQCYVVAACGCRHHEKRASYGSHSVVDPMGTVVVARCSGPGICLARIIDLNTL 297

OY 337 ROLRRHLVPFOHRRPDLVGNLGHPLS 362
      |||||||
Db 298 ROLRRHLVPFOHRRPDLVGNLGHPLS 323

RESULT 4
O9DBF2 PRELIMINARY: PRT: 290 AA.
ID O9DBF2;
AC O9DBF2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE 1.
```

GN NIT1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RX MEDLINE=21085660; PubMed=11217651;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Cusack-Stinch S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK004988; BAB23723.1;
 DR MGD: MGI:1350916; Nf1l.
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 SQ SEQUENCE 290 AA; 31886 MW; 34493DBFF2170C71 CRC64;
 Query Match 71.48; Score 1372; DB 11; Length 290;
 Best Local Similarity 87.5%; Pred. No. 3.6e-114;
 Matches 252; Conservative 22; Mismatches 14; Indels 0; Gaps 0;
 QY 75 SSSCEPLVAVCOVTPKQONFKTCALVREARLACLAFLPAPFPIARDPETHL 134
 DB 3 SSTSWEPLVAVCOVTPKQONFKTCALVREARLACLAFLPAPFPIARDPETHL 62
 QY 135 HISEPLGLKLEETYLARCGIWLISGFHERGQDWEOTOKIYNCHVLLNSKGAIVATY 194
 DB 63 LIEPLGLDGLGYSQLARCGIWLISGFHERGQDWEONQKTYNCHVLLNSKGSVASY 122
 QY 195 RKTHLCDVEIPGGPMPCESTNMPGSLSPVSTPACRTGLAVCYDMKFPPLSLALQAG 254
 DB 123 RKTHLCDVEIPGGPMPCESTNMPGSLSPVSTPACRTGLAVCYDMKFPPLSLALQAG 182
 QY 255 AEITTPSASGTGPAHMEVLLRARAIEIQCYYVAAACGRHHEKRSYGHSHVVDPMG 314
 DB 183 AEITTPSASGTGPAHMEVLLRARAIEIQCYYVAAACGRHHEKRSYGHSHVVDPMG 242
 QY 315 TVVARCEGPGCLARIDLNYLRRLRHLRVFQHRRPDLVGSGLHPLS 362
 DB 243 TVVARCEGPGCLARIDLNYLRRLRHLRVFQHRRPDLVGSGLHPLS 290
 RESULT 5
 ID 091860 PRELIMINARY; PRT; 288 AA.
 AC 091860;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE NIT1 PROTEIN 1.
 GN NIT1.
 OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=6355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC PubMed=10959838;
 RX Pace H.C., Hodawadkar S.C., Draganescu A., Huang J., Bieganski P.,
 RA Pekarsky Y., Croce C.M., Brenner C.;
 RT "Crystal structure of the worm NITfhl1 Rosetta Stone protein reveals a
 RT NIT tetramer binding two Fhl dimers.";
 RL Curr. Biol. 10:907-917(2000).
 DR EMBL: AF284575; AAF87104.1;
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 SQ SEQUENCE 288 AA; 32061 MW; B3673A80609BF042 CRC64;
 Query Match 49.88; Score 956; DB 13; Length 288;
 Best Local Similarity 62.7%; Pred. No. 3.7e-77;
 Matches 173; Conservative 44; Mismatches 59; Indels 0; Gaps 0;
 QY 82 PLVAVCOVTPKQONFKTCALVREARLACLAFLPAPFPIARDPETHLSEPIG 141
 DB 7 PLVAVCOVTPKQONFKTCALVREARLACLAFLPAPFPIARDPETHLSEPIG 66
 QY 142 GLTLEETYLARECGIWLISGFHERGQDWEOTOKIYNCHVLLNSKGAIVATYRTHLCD 201
 DB 67 GDTIORTYLARECGIWLISGFHERGQDWEOTOKIYNCHVLLNSKGAIVATYRTHLCD 126
 QY 202 VEIRGQGPCESTNMPGSLSPVSTPACRTGLAVCYDMKFPPLSLALQAGAILTYP 261
 DB 127 VDLQNGVSLRSESTLPGALIRPITSPAGRTGLAVCYDMKFPPLSLALQAGAILTYP 186
 QY 262 SAFGSIPTPAHMEVLLRARAIEIQCYYVAAACGRHHEKRSYGHSHVVDPMGVARCS 321
 DB 187 SAFGSIPTPAHMEVLLRARAIEIQCYYVAAACGRHHEKRSYGHSHVVDPMGVARCS 246
 QY 322 EGPGLCLARIDLNYLRRLRHLRVFQHRRPDLVGNL 357
 DB 247 EGPGLCLARIDLNYLRRLRHLRVFQHRRPDLVGNL 282
 RESULT 6
 ID 076464 PRELIMINARY; PRT; 460 AA.
 AC 076464;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE NITFHL PROTEIN.
 GN NITFHL OR CG7067.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
 RA Adair J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Houston D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwan C.,
RA Jastli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Put V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sieden Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE-98337986; PubMed-9671749;
RA Pekarisky Y., Campigillo M., Siprasvill Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhl homologs are encoded as fusion proteins in
RT *Drosophila melanogaster* and *Caenorhabditis elegans*.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749 (1998).
RN EMBL: AE003467; AAF4347.1; -;
DR EMBL: AF069989; AAC39137.1; -;
DR HSSP: P49789; 2FTF.
DR FLYBASE: FBgn0024945; NitFhl.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001310; HIT.
DR Pfam: PF00795; CN_hydrolase; 1.
DR Pfam: PF01230; HIT; 1.
DR PROSITE: PS00892; HIT; 1.
SQ SEQUENCE 460 AA; 52231 MW; 81121A00BC37706 CRC64;

Query Match 37.0%; Score 710.5; DB 5; Length 460;
Best Local Similarity 48.6%; Pred. No. 4,5e-55;
Matches 143; Conservative 47; Mismatches 97; Indels 7; Gaps 3;

OY 62 VACADPRPAMISSSSCLPLVAVCOVSTPDKOONFKTCALVREARLGAFLPDE 121
DB 13 VAIHQQLERMSVQKRKDSATLAVGOMRSTSDKANLSQVLETVLRARSQNMCLFDE 72
OY 122 ADFDIARDAETLHSEPLGKLLSEYTLARCEGLMSLGGFHERGOWEOTQKLYNCH 181
DB 73 CCDFVEESTQTITELSEGDLGELMAYRELAKCNKWTISLGGYHERND-----QKLFNAH 127
OY 182 VLNSKGAAVAVATYRKTLDVETPGGPMCSNSTMPGSPLESVPSTPAKGTGLAVCYDM 241
DB 128 VILNKGELAAVYRKLMHEDVTTK-EVRLRESDTVPKGCLERPVSTPGVIGLQICYDL 186
OY 242 REPPELSLAQAQAEILTPSAFSGSTTGAHMEVLLRAAIEITQCIVVAAOCGRHNEKR 301
DB 187 RAEPRAVILRLKGANLITPSAFYATYAKAHWEILLRAAIEITQCIVVAAOCGRHNEKR 246
OY 302 ASYGSMMVDPGCTVAVARCEGP-GICLARIDNLRLPVOHRRPDLX 354
DB 247 OSWGHSMIVSPMGNVLAOCSEBOLDIGTAEVDLSVLQSLQYTPCFEHRNDIY 300

RESULT 7
ID 076463 PRELIMINARY; PRT; 440 AA.
AC 076463;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRILASE AND FRAGILE HISTIDINE TRIAD FUSION PROTEIN NITFHL.
GN NITFHL OR Y56A3A.13.
OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; *Caenorhabditis*.
OX NCBI_TaxID=6239;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE-98337986; PubMed-9671749;
RA Pekarisky Y., Campigillo M., Siprasvill Z., Druck T., Sedkov Y.,
RA Tilib S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,
RA Buchberg A.M., Mazo A., Brenner C., Croce C.M.;
RT "Nitrilase and Fhl homologs are encoded as fusion proteins in
RT *Drosophila melanogaster* and *Caenorhabditis elegans*.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99069613; PubMed-9851916;
RA none;
RT "Genome sequence of the nematode *C. elegans*: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
RN EMBL: AF069986; AAC39136.1; -;
DR EMBL: AL132860; CAB60517.1; -;
DR HSSP: P49789; 2FTF.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001310; HIT.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR Pfam: PF01230; HIT; 1.
DR PROSITE: PS00892; HIT; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 440 AA; 49936 MW; 34EBD223CE92975 CRC64;

Query Match 36.5%; Score 702; DB 5; Length 440;
Best Local Similarity 50.2%; Pred. No. 2,4e-54;
Matches 136; Conservative 43; Mismatches 90; Indels 2; Gaps 1;

OY 84 VAVCOVSTPDKOONFKTCALVREARLGAFLPDEAFARDAETLHSEPLGK 143
DB 17 IAVCOVSTPDKOONFKTCALVREARLGAFLPDEAFARDAETLHSEPLGK 76
OY 144 LEEYTLARECGVLSLGGFHERGOWEOTQKLYNCHVLSKGAVVATYRKTLDV 203
DB 77 YMEKYRELARKNHVITSLGGLHNK--DPSDAHPWNTHTIIDSDVTRAEVYNKLHFDLE 134
OY 204 IPGQPMCSNSTMPGSPLESVPSTPAKGTGLAVCYDMREPPELSLAQAQAEILTPSA 263
DB 135 IPGKVALMESEFSKATEMIPIVDPITIGRIGLSTICDVAFPELSLNNRRGQDLSPSA 194
OY 264 FGSITPAHMEVLLRAAIEITQCIVVAAOCGRHNEKRASYGSMMVDPGCTVAVARCEG 323
DB 195 FTNLNGLAMEVLLRAAIEITQCIVVAAOCGRHNEKRASYGSMMVDPGCTVAVARCEG 254
OY 324 PGICLARIDNLRLPVOHRRPDLX 354
DB 255 VDMCFEALDLSVDTLREMPVFSHRRSDLY 285

RESULT 8

09LE50
ID 09LE50 PRELIMINARY: PRT: 316 AA.
AC 09LE50:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NITRILASE 1 LIKE PROTEIN.
GN ATAG08790.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA McCombie R.W., Spiegel L.A., Huang E.N., Nascimento L.U.,
RA de la Bastide M., Vill D.M., Preston R.R., Matero A., Shah R.,
RA O'Shaughnessy A., Rodriguez M., Shekher M., Schutz K., See L.H.,
RA Swaby I., Habermann K., Dedhia N.N., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vill D.M.,
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161813; CAB82115.1; -
DR EMBL; AL161512; CAB78004.1; -
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
SQ SEQUENCE 316 AA: 34807 MW: 1D740F48DF9A202C CRC64;

Query Match 33.6%; Score 645.5; DB 10; Length 316;
Best Local Similarity 44.2%; Pred. No. 1.7e-49;
Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;

QY 68 RRRRAAIISSSCGELP-----VAVCOVSTPDKQONFKTCAELVREARIGACL 116
DB 12 KSLFTTRITLSQIPLTMATVTKTVRAAQMSTVNDLMTNFATCSRLVQEAALAGAKL 71
QY 117 AFLPEAFDFIARDPAETHTLSEPLGKLEEYTLARECGMLSLGSPFERGQDWGQTK 176
DB 72 ICFPENFSVGKKEGESVKAIEPLDGPVWERICSLARDNSINIMLSLGGFQDERFDD----TH 127
QY 177 IYNCHVLNLSKGAVVATYKTHLCDEIPLGOGPCESSNTMPGPSLESPVPAGKIGLA 236
DB 128 LCNFTVHVIDDAMGIRDTYQKMLFDVDPGSSYKSSSFVPTKIVS--VDSVGRGLGL 186
QY 237 VCYDMRFPFLSLALA-QAAGELITYSAGSITGPRAHMEYLLRAKALETCQYVAAAQCG 295
DB 187 VCYDLRFPKTIYQOLRFEORAQVLLVPSAFTKYVGEAHMETLLRAKALETCQYVAAAQAG 246
QY 296 RHHRKRASGSHMVDVPGTVAARCE--GPGICLARIDLNTYLRRLPV 345
DB 247 KNEKRRESIGDTLLIDPMGTVYGRPLDRVSTGIIVADIDFSLIDSVRTKMPI 298

RESULT 9
AC 094660 PRELIMINARY: PRT: 276 AA.
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHEICAL 30.4 KDA NITRILASE-LIKE PROTEIN C651.02 IN CHROMOSOME II.
GN SPBC651.02.

OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-972;
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE N-TERMINAL OF NITRILASES.
DR EMBL; AL035570; CAB37598.1; -
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
KW Hypothetical protein; Lyase.
FT ACT_SITE 156
SQ SEQUENCE 276 AA: 30421 MW: 07AA741A54297E07 CRC64;

Query Match 30.0%; Score 575.5; DB 3; Length 276;
Best Local Similarity 44.2%; Pred. No. 2.5e-43;
Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;

QY 81 LPLVAVCOVSTPDKQONFKTCAELVREARIGACLPEAFDFIARDPAETLHT-SEP 139
DB 1 MTLAAVAGQLNSSGSLIKNLALICKELISQAARAKACIFFPEADDFIANDENAEILTNHP 60
QY 140 LGGKLEEYTLARECGMLSLGSPFERGQDWGQTKYINCHVLNLSK-----GAVVA 192
DB 61 DCSKTRFDVRESATKSHIFVNI--CVHE-----PSVKKN--KLNSLFIPELHGEEIS 110
QY 193 TYKTHLCDEIPLGOGPCESSNTMPGPSLESPVPAGKIGLAVCYDMRFPFLSLALAQ 252
DB 111 RYSKAHLFDVEINNGPTLKESNTLNGEALLPCKKPLGKVGSAICDIFPPQAIKRLN 170
QY 253 AGAEILTYPSAFSGITGPRAHMEYLLRAKALETCQYVAAAQCGRHHRKRASGSHMVDP 312
DB 171 MGAHITTPSAFTKTEKGAHMEYLLRAKALDSQCYVAPAGQGHNRKRASGSHMIVDP 230
QY 313 MGTVVAARCE--GP-GICLARIDLNTYLRRLPVFQHRRPDLX 354
DB 231 MGVVIAQVSDISSPNGLIFADLDLNLVDVRYTYIPLL--RRNDLY 273

RESULT 10
ID 027839 PRELIMINARY: PRT: 272 AA.
AC 027839:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE N-CARBAMOYL-D-AMINO ACID AMIDOHYDROLASE.
GN MTH1811.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-DELTA H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.T., Mao J.-T., Rice P., Nolling J., Reeve J.N.;
RT Complete genome sequence of Methanobacterium thermoautotrophicum
RT delta: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000934; AAB86277.1; -
DR InterPro: IPR003010; CN_hydrolase.

Db 224 VVPMGCVILAKAGTEEAIVYSDIDLKTLAEIRQIIPVFRQKRSDLY 269

RESULT 13
ID 083040 PRELIMINARY; PRT; 270 AA.
AC 083040;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF270 PROTEIN.
OS Plectonema boryanum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
OX NCBI_TaxID=1184;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D65;
RA Matsumura T., Fujita Y., Hase T.;
RT "Cyanobacterial ferredoxin gene."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB01194; BA32602.1;
DR InterPro: IPR001110; UPF0012.
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
SQ SEQUENCE 270 AA; 29639 MW; 131C447C00BF740 CRC64;

Query Match 24.9%; Score 479; DB 2; Length 270;
Best Local Similarity 39.6%; Pred. No. 9.3e-35;
Matches 112; Conservative 40; Mismatches 95; Indels 36; Gaps 7;

QY 83 LVAVQVTSPPKQNFKAELVREARLGAFLPFAFDPAETLHSEPLG 142
DB 5 LAAVQVNSPDLQKLAQAELEIIVRRGAEILGIPENFSFGDE-----G 53
QY 143 KLEEVYQLREGSLM-----LGFGH-----ERQDMQOTQKIVNCHVLLNSKG 188
DB 54 KIRQA-DEIVVESKEFLKTAQARQVYVILGGYVPEVPG-----KVNTALLVAPNG 105
QY 189 AVVATYKTHLCVDEIPGQPMCESNSTMPSPLESFV--STPAKIGLAVCYDMRPPEL 246
DB 106 ELARVYKHLFDVNLDPGNTYRESATVAVSTR-PRIPHSKDGLNIGLSYCVRRPEL 164
QY 247 SLAQAQAEILYPSAFSGSTTGPAHWEVLARAFETQCYVVAACGRHNEKRAYGH 306
DB 165 YRQMTQGAEVLFPAPAFHTGKDHQVLLQARAIENTCYVIAPQOTGRHNSLSRSHGH 224
QY 307 SMVVDPMGTAVVARGSEGPLARIDNLYRLQLRRHLVPEQHR 349
DB 225 AMIDPMGVILADAGMTGVAIAETAPRLERQVRROMPSLQHR 267

RESULT 14
ID 09JHM2 PRELIMINARY; PRT; 276 AA.
AC 09JHM2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NIT PROTEIN 2 (1190017B19RIK PROTEIN).
OS NIT2 OR 1190017B19RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX PdbMed=10959838;
RA Pace H.C., Hodawadekar S.C., Draganescu A., Huang J., Bleganowski P.,
RA Pekarsky Y., Croce C.M., Brenner C.;
RT "Crystal structure of the worm NitFhit Rosetta Stone protein reveals a

RT Nit tetramer binding two Fhit dimers.";
RL Curr. Biol. 10:907-917(2000).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Botfield D., Boujunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohitsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AF284573; AB87102.1;
DR EMBL: AK004535; BAB23354.1;
DR MGD: MGI:1913477; 1190017B19RIK.
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
SQ SEQUENCE 276 AA; 30501 MW; 740FDC44978326D6 CRC64;

Query Match 24.7%; Score 474.5; DB 11; Length 276;
Best Local Similarity 37.2%; Pred. No. 2.4e-34;
Matches 103; Conservative 52; Mismatches 103; Indels 19; Gaps 6;

QY 84 VAVQVTSPPKQNFKAELVREARLGAFLPFAFDPAETLHSEPL 140
DB 6 LALIQLVSSIKSDNLTFRACSLVREARAGANIVSLPCFN----SPGTGYFPDYAKI 61
QY 141 GKLLEEVYQLREGSLM-----LGFGH-----ERQDMQOTQKIVNCHVLLNSKGA VAVATYKTHLC 200
DB 62 PESTQKLEEVAKESSTYLGSGIPE-----EDAGKLYNTCSVGPDSGLYKRRKTHLF 116
QY 201 DVEIPGQPMCESNSTMPSPLESFVSTPAKIGLAVCYDMRPPELSTALAQAQAEILTY 260
DB 117 DIDVPGKITTFQESKTLSPGDSF-STFDTPYCKVAGIGICYDMRFAELAOIYARQGGOLLVY 175
QY 261 PSAFSGSTGPAHWEVLARAFETQCYVVAACGRHNEKRAY--CHSNVVDPMGTAVV 317
DB 176 PGAFNLTGTPAPAWELQRAAVDQVYVATASPA---RDDRASVYAWGHSVVDPMGCVL 232
QY 318 ARCSGPGELARIDNLYRLQLRRHLVPEQHRPPDY 354
DB 233 TRAGTEETILYSDIDLKTLAEIRQIIPILKORRADLY 269

RESULT 15
ID 09KUD4 PRELIMINARY; PRT; 275 AA.
AC 09KUD4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN VC0421.
GN VC0421.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae";
 RL Nature 406:477-483(2000).
 DR EMBL; AE004129; AAF93594.1; .
 DR TIGR; VC0421; .
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE; PS01227; UPF0012; 1.
 KW Complete proteome.
 SQ SEQUENCE 275 AA; 30630 MW; F1342B0BF519EE4 CRC64;

Query Match 22.7%; Score 436.5; DB 2; Length 275;
 Best Local Similarity 38.1%; Pred. No. 5.7e-31;
 Matches 104; Conservative 41; Mismatches 111; Indels 17; Gaps 5;

```

OY 64 VAVCOVTSPPDKONFRTCAELVREARLGACLAFLPEAFDFTA----RDPAETLHSE 138
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 4 VGLVOMTSGPEISRNLTYLKEQVAKLAQOGAOMIVTPENALLIGNRQYHQAEESLD--- 60
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 139 PLGGKLELEYTOLARECGLWLSLGFHERGQDWEOTQKIYNCHVLNLSKGAVVATYRKT 198
   | : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 61 --HGPVQHALLASLAKKEYGVLLIGSMPIRHAEGVTTSSL-----LWNAQGERVAVYDKL 113
   | : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
OY 199 LCQVEI--PGGPRICESNSTMPGSPLESVPSTPAKTIGLAVCYDMRPPELSLAQAQAEI 257
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 114 MFDVDVADGHQRYRESEFTTPGGQV--VVTSTPGALGLSTICYDVRPHLYADLRROGAOI 172
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 258 LFTPSAFSGITGPAHMEVLLRARALETQCYVVAQAOCGRHHEKRAKSYGHSWVDPMGTVV 317
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 173 LTVPAAFYANTVGAHMEVLLRARALETQCYVVAQAOCGRHHEKRAKSYGHSWVDPMGTVV 317
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 318 ARCSEGGICLARIIDLNYLRQLRRHLPVQHR 350
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 233 ANLGAEVQSKVVEEDLATLDSVRRAMPITQHT 265
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Search completed: April 26, 2002, 17:02:23
 Job time: 461 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 17:02:50 ; Search time 21.41 Seconds

(without alignments)
619.928 Million cell updates/sec

Title: US-09-357-675C-25

Sequence: 1 PLAAAXLAPDRPRDRLRMV.....LPVFQHRPRDLGNLGHPLS 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_39;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	519	27.0	307	1 YJM6_YEAST	P47016 saccharomyc
2	490	25.5	372	1 Y601_SYNY3	P55175 synchocyst
3	448.5	23.3	322	1 YAU8_SCHPO	Q10166 schizosacch
4	420	21.9	291	1 YI85_YEAST	P49954 saccharomyc
5	356	18.5	262	1 YBEK_ECOLI	P39874 escherichia
6	353	18.4	262	1 YBEK_ECO57	P58054 escherichia
7	262.5	13.7	340	1 Y480_MYCTU	Q11446 pseudomonas
8	254.5	13.2	285	1 YPOQ_PSEFL	P55176 pseudomonas
9	231	12.0	234	1 YAG5_STALU	P55178 staphylococ
10	221.5	11.5	346	1 NRJ3_ARATH	P46010 arabidopsis
11	220.5	11.5	261	1 YAG5_STRAU	P55177 staphylococ
12	214.5	11.2	346	1 NRJ1_ARATH	P32961 arabidopsis
13	198.5	10.3	339	1 NRJ2_ARATH	P32962 arabidopsis
14	195	10.2	346	1 ALAM_PSEAE	P14436 pseudomonas
15	179.5	9.3	393	1 BUP_RAT	Q03248 rattus norv
16	178.5	9.3	355	1 NRJ4_ARATH	P46011 arabidopsis
17	159	8.3	349	1 NRJ4_TOBAC	Q42965 nicotiana t
18	155	8.1	513	1 YHCK_BACSI	P54608 bacillus t
19	142	7.4	256	1 YAFV_ECOLI	Q47679 escherichia
20	142	7.4	345	1 ALAM_RHOER	Q01360 rhodococcus
21	131	6.8	576	1 NAE2_THEMA	Q90400 thermotoga
22	115.5	6.0	199	1 YI04_YEAST	P40447 saccharomyc
23	114.5	6.0	679	1 NAE2_MYCTU	P71911 mycobacteri
24	113.5	5.9	356	1 NRJ1_ALCEA	P20960 alcalligenes
25	112.5	5.9	368	1 CYRH_GLOSO	P32964 glaucocercos
26	108.5	5.6	552	1 NAE2_RHOCA	Q03638 rhodobacter
27	102	5.3	383	1 NRJ1_RHOER	Q02068 rhodococcus
28	102	5.3	567	1 NAE2_AOUAE	Q67091 aquifex aeo
29	100	5.2	365	1 NRJ2_RHOER	Q03217 rhodococcus
30	98.5	5.1	511	1 LNT_PSEAE	Q92186 pseudomonas
31	97.5	5.1	680	1 NAE2_MYCLE	Q9cb66 homo sapien
32	96	5.0	794	1 SEIL_HUMAN	Q9ubv2 homo sapien
33	94.5	4.9	433	1 FUR4_HUMAN	Q62994 rattus norv

34	94	4.9	542	1 LNT_CHLTR	O84539 chlamydia t
35	93	4.8	349	1 NRJ6_KLEPO	P10045 klebsiella
36	93	4.8	519	1 LNT_SYNY3	P74055 synchocyst
37	92.5	4.8	520	1 VNN2_HUMAN	Q95498 homo sapien
38	91.5	4.7	541	1 LNT_RHIME	Q52910 rhizobium m
39	91	4.7	512	1 VNN1_MOUSE	Q92048 mus musculu
40	90	4.7	439	1 LNT_AOUAE	Q67000 aquifex aeo
41	90	4.7	2209	1 POU3_POLIS	P03301 poliovirus
42	89.5	4.7	642	1 HEM1_RAT	P13195 rattus norv
43	89.5	4.7	1257	1 ERB2_RAT	P06454 rattus norv
44	89	4.6	1255	1 ERB2_HUMAN	P04626 homo sapien
45	88	4.6	790	1 SEIL_MOUSE	Q92296 mus musculu

ALIGNMENTS

RESULT	ID	YJM6_YEAST	STANDARD	PRT	307 AA.
AC	P47016;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DE	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	HYPOTHETICAL 34.7 KDA PROTEIN IN SPT10-GCD14 INTERGENIC REGION.				
CN	YJL126W OR J0706.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=5288C / FY1679;				
RA	MEDLINE=97103775; PubMed=8948101;				
RT	Ciepluch C., Kordes E., Pujol A., Jauniaux J.-C.;				
RT	"Sequencing analysis of a 40.2 kb fragment of yeast chromosome X				
RT	reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,				
RT	SPT10, GCD14, RPEL, PHO86, NCA3, ASF1, CCF7, GZF3, two tRNA genes,				
RT	three remnant delta elements and a Ty4 transposon."				
RL	Yeast 12:1471-1474(1996).				
CC	- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: 249401; CA89421.1; -.				
DR	SGD: S0003662; YJL126W.				
DR	InterPro: IPR003010; CN_hydrolase.				
DR	InterPro: IPR001110; UPF0012.				
DR	Pfam: PF00795; CN_hydrolase; 1.				
DR	PROSITE: PS01227; UPF0012.1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 307 AA; 34693 MW; 48787C43B10A828E CRC64;				
Query Match	27.0%; Score 519; DB 1; Length 307;				
Best Local Similarity	37.6%; Pred. No. 6.9e-38;				
Matches 114; Conservative 56; Mismatches 105; Indels 28; Gaps 6;					
QY	80 ELPLVAVCQYVTPDKQNFCTCAELVREARLGCACLAFLPEARDFTARDPAETIHLT---				136
DB	4 KLRKVAVAQQLSSADLRLKNLVKELISEAQKADAVFLPEASDYLVSQNPILHSRYLAOK				63
QY	137 SEPLGKLEBYTOLARECGILWSLG-GFH--ERGGQ-WEGTQKIYMCVHLNSKGAIVA				192
DB	64 SPKRTIOLQSSITDLVBDNSRNDIVSIGVHLPSPEODLLEQNDVRVNVLLYIDHEGKITQ				123
QY	193 TYRRTIHCADVEIIPQGMCSNSTMPPSLSEPVSTPAGKIGLAVCVDMRPPELSLALAQ				252

DB 124 EKQKHLFEVDVDPNGPILKESKSVOPGKAIPIPIESPLKLSAICQDIRPEFFSLKLS 183
OY 253 AGAEILTPSAGSITGPALHWEILRLARAIETQCYVVAAGCRH----- 297
DB 184 MGAELICFPASATITKTGEHWEILGRARAVDTQCYVLMGQGMHDLSPENKOSHMSA 243
OY 298 ---HEKRASYSYSHMVDPNGTVVARC---SEGPGLCLARIDLYLRLRRLHVPFQHRP 351
DB 244 LKSSRSRESGSHSVIDPAGKIIAHADPSTVGQILADLDRELRLQEIIRKPKLIMNQRRD 303
OY 352 DLX 354
DB 304 DLF 306
RESULT 2
Y601_SYNY3 STANDARD; PRT; 272 AA.
ID Y601_SYNY3
AC P55175;
DT 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 30.2 KDA PROTEIN SLL0601.
GN SLL0601.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL: D64002; BAA10370.1;
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 272 AA; 30191 MW; 7787BA9B2BE003A1 CRC64;

Query Match 25.5%; Score 490; DB 1; Length 272;
Best Local Similarity 40.2%; Pred. No. 2e-35;
Matches 111; Conservative 46; Mismatches 99; Indels 20; Gaps 7;

DB 33 LVAVQCVSTPRKQONFKTCAELVREARLGLACLAFLPEAFDFIARDPAETLHLSEPLG- 141
OY 5 LAAALQMTSRPLNTLELAEELIDLAIVROGAELVGLPNEFAFLG---ETEKLEQATAI 61
DB 142 GKLEETPD-LAREGGLMELSGF-----HERGQDEQFOKLYNCHVILLNSKGAVAVATVR 195
OY 52 AATATEKFLDTMAQRQVITLAGFPPEPVAGEAG-----KATNTATLTPMGQELARKH 114
DB 136 KTHLCLDVEIPGCGPMCESNSTMPGSLSPV---STPAGKIGLAVCYDMRFPETSLALAOA 253
DB 115 KHLFDVNVNPDGNTYWESTVMAQGY-PPVYHSDSFGMLGLSICDVAFPELXYRLSLQ 173
OY 254 GAEILTPSAGSITGPALHWEILRLARAIETQCYVVAAGCRHHEKRASYSYSHMVDPW 313

DB 174 GADVLEVPAAFLAYTKCKDMQVLLQARAIENTCYVLAPOQTSCHYERRHTGHAMIDPW 233
OY 314 GTVVAARCSEPGICLARIDLYLRLRRLHVPFQHR 349
DB 234 GVLADAGEKPGLAIAETINPDRILKQVROOMPSLOHR 269
RESULT 3
YAU8_SCHPO STANDARD; PRT; 322 AA.
ID YAU8_SCHPO
AC Q10166;
DT 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 35.7 KDA PROTEIN C26A3.11 IN CHROMOSOME I.
GN SPAC26A3.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RA Mclean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL: Z69240; GAA9334.1;
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR001110; UPF0012.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS01227; UPF0012; 1.
KW Hypothetical protein.
SQ SEQUENCE 322 AA; 35679 MW; EA6F39B160C7F49F CRC64;

Query Match 23.3%; Score 448.5; DB 1; Length 322;
Best Local Similarity 32.8%; Pred. No. 9.8e-32;
Matches 108; Conservative 63; Mismatches 123; Indels 35; Gaps 9;

DB 46 RLFLSLCPGLR--IFOLSVLCAQPPRRAMAISSSCCELP-----LVAVQVSTPRDKQ 97
DB 4 KRFGLVQKTRSFPSLNCYTR---NIMSVASSL-VKDRPARRIGLYOLANKDSE 59
OY 98 NFKTCAELVREARLGLACLAFLPEAFDFIARDP-----AETLHLSEPLGKLEET 148
DB 60 NQQLRLKYLEAKKGSNVIPLPIFN---SPYGTGFNOYAEPLEESSP-----SYQAL 111
DB 149 TOLARECGMLSLSGFHGHGQDEQFOKLYNCHVILLNSKGAVAVATVR 208
OY 112 SSMARDQTKYTLFGGSIPER-----KDGKLYNTAMFDPGSKLIANHRKHLHLDIDIPGV 166
DB 209 PMCESNSTMPGSLSPVSTPAGKIGLAVCYDMRFPETSLALQAGAEILTPSAGSIT 268
DB 167 SFRSDSLSPGAM-TMVDTEYKFGGLGICVDIRPELAMIAARNGCSVMYIPGAFNLST 225
OY 269 GRAHWEILRLARAIETQCYVVAAGCRHHEKRASYSYSHMVDPNGTVVARCSEPGICL 328
DB 226 GLHWEILRLARAVNVNEMVACCAPARDMADYHSHGSHSTVVDPEGKVIATTDKPSIV 285
OY 329 ARTIDNTYLRQRLRHLPVFOHRRPDLXGNL 357
DB 286 ADIDPSVMSTARNSVPIYTORRFDVSEV 314

```

RESULT 4
ID YBEM_ECOLI STANDARD; PRT; 262 AA.
AC P39874;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHEMETICAL PROTEIN YBEM.
GN YBEM OR B0625/B0626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horituchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Nemat A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-51 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95075315; PubMed=7984109;
RA Yamana K., Mitani T., Ogura T., Niki H., Hiraga S.;
RT "Cloning, sequencing, and characterization of multicopy suppressors
RT of a mxb mutation in Escherichia coli."
RL Mol. Microbiol. 13:301-312(1994).
RN [5]
RP SEQUENCE OF 31-78 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Robison K., O'Keefe T., Church G.M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [6]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JAN-2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC POSITION 67.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL: U9102; AAB67751.1; -.
CC SGD: S0004343; YLR351C.
CC InterPro: IPR003010; CN_hydrolase.
CC InterPro: IPR001110; UPF0012.
CC Pfam: PF00795; CN_hydrolase; 1.
CC PROSITE: PS01227; UPF0012; 1.
CC KMW Hypothetical protein.
SQ SEQUENCE 291 AA; 32549 MW; A81374412008827 CRC64;

```

```

Query Match 21.9%; Score 420; DB 1; Length 291;
Best local Similarity 34.5%; Pred. No. 2,6e-29;
Matches 98; Conservative 53; Mismatches 115; Indels 18; Gaps 7;

```

```

DB 84 VAVCQVT-STPDKQONFKTCAELVREAR--LGACIAFLDEAFD-----FIARDPAET 133
DB 13 VALVQLSSSPDKMANLQRAATFERAMKEQDPTKLVLEDECFNSPYSTDF--RKYSSEV 70
OY 134 LHTSEPLGKLLDEYQLANECGLMSLGGFHEHGQDMEDQTKYNNCHVLNSGAVYAT 193
DB 71 INKEP--STSVQFLSMIAKFKILLGGTPELD---PRTDKYNTSIIFNEGKLIK 125
OY 194 YRRTHLCDVEIPCGPSCNSNTPWGPSPVSTPAKTIAGLAVCYMRPELSIALQA 253
DB 126 HRRVHLFDVDPINGISFHESETLSPEEK-STTIDTKYKGGVGYCYDMRPELMLARK 184
OY 254 GAELIYPSAFSGSITGPAHVEVLIRARAIEQCYVAAACGRRHEKRAYSGHSMVDPW 313
DB 185 GARFAMITPSAFNTVTGGLHMLLARSRAVDNQYVMLCSPARNLQSSYHAHGSHIVDPR 244
OY 314 GTVVARCSEGPGLARIDLVYLRQLRHLRYPQHRRLDLYGNL 357
DB 245 GKIVAAGEEETIYAELEDEVIESFRQAVPLTKQRFPDYSDV 288

```

```

RESULT 5
YBEM_ECOLI

```

```

ID YBEM_ECOLI STANDARD; PRT; 262 AA.
AC P39874;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHEMETICAL PROTEIN YBEM.
GN YBEM OR B0625/B0626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horituchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Nemat A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-51 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95075315; PubMed=7984109;
RA Yamana K., Mitani T., Ogura T., Niki H., Hiraga S.;
RT "Cloning, sequencing, and characterization of multicopy suppressors
RT of a mxb mutation in Escherichia coli."
RL Mol. Microbiol. 13:301-312(1994).
RN [5]
RP SEQUENCE OF 31-78 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Robison K., O'Keefe T., Church G.M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [6]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JAN-2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE K12 SEQUENCE
CC DIFFERS FROM THE WILD-TYPE SEQUENCE DUE TO A FRAMESHIFT IN
CC POSITION 67.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL: AE000167; AAC73726.1; ALT_FRAME.
CC EMBL: AE000167; AAC73727.1; ALT_FRAME.
CC EMBL: D90703; BAA35268.1; ALT_FRAME.

```

DR EMBL: D90703; BAA35269.1; ALT_FRAME.
 DR EMBL: U82598; AAB40826.1; ALT_FRAME.
 DR EMBL: D28497; ; NOT_ANNOTATED_CDS.
 DR EMBL: U38811; AAA81032.1; ALT_FRAME.
 DR Ecogene: EG3162; ybeM.
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 262 AA; 28625 MW; 116B11439F83EAD2 CRC64;

Query Match 18.5%; Score 356; DB 1; Length 262;
 Best Local Similarity 35.8%; Pred. No. 8.3e-24;
 Matches 100; Conservative 42; Mismatches 101; Indels 36; Gaps 11;

QY 83 LVAVCQVTSPPDKQNFKTCALVREARAGACLAFLPEAFDIARDPAE---TLHLSEP 139
 DB 2 LVAGGPAVTSWEKNAEICASLMAQAENDASLFALEPA--LLARDHDADLSVKS AOL 59
 QY 140 LGSKLLEEYVTLAREC--GLMLSLGFFH---ERGDMEQOTOKIYNCHVLNSKGAVVATY 194
 DB 60 LEEFPL---GRLRRESKRNMTTILITIHVSTPGRAM-----NMLVALQA-GNIVARY 108
 QY 195 RKTHLCDVEIPGGPMCESNSTMPGSPLESVPSTPAKIGLAVCYDMRPELSLALAAQ 254
 DB 109 AKHLVDA-----FAIOESRRVDAGNEIAPVLEVEGKMKVGLMTCYDLRPELALAAQALOG 163
 QY 255 AELLTPSAFGSTTGPRA---HMEVLLRARALETQCYVVAAGCGRHHEKASGHSWVD 311
 DB 164 AELLTPSAFW--VRGPKHEHMTSLLAARALDITTCYVVAAGCGC---NKNIGOSRTID 216
 QY 312 PMGTVAACSEGGPLCLARIDLNYLRQLRHLRPFQHR 350
 DB 217 PFGVTIAASEMPALLMAEVTPEPVQVRAQLVPLNRR 255

RESULT 6
 YBEM_ECO57
 ID YBEM_ECO57 STANDARD: PRT; 262 AA.
 AC P58054;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN YBEM.
 GN YBEM OR 20771 OR EC50664.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=8334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RA MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RA MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, on the basis
 CC of the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AE005242; AAG54960.1; ALT_INIT.
 DR EMBL: AP002352; BAB34087.1; .
 DR PROSITE: PS01227; UPF0012; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 262 AA; 28681 MW; 74CF04FD3B9CCIDF CRC64;

Query Match 18.4%; Score 353; DB 1; Length 262;
 Best Local Similarity 35.5%; Pred. No. 1.5e-23;
 Matches 99; Conservative 42; Mismatches 102; Indels 36; Gaps 11;

QY 83 LVAVCQVTSPPDKQNFKTCALVREARAGACLAFLPEAFDIARDPAE---TLHLSEP 139
 DB 2 LVAGGPAVTSWEKNAEICASLMAQAENDASLFALEPA--LLARDHDADLSVKS AOL 59
 QY 140 LGSKLLEEYVTLAREC--GLMLSLGFFH---ERGDMEQOTOKIYNCHVLNSKGAVVATY 194
 DB 60 LEEFPL---GRLRRESKRNMTTILITIHVSTPGRAM-----NMLVALQA-GNIVARY 108
 QY 195 RKTHLCDVEIPGGPMCESNSTMPGSPLESVPSTPAKIGLAVCYDMRPELSLALAAQ 254
 DB 109 AKHLVDA-----FAIOESRRVDAGNEIAPVLEVEGKMKVGLMTCYDLRPELALAAQALOG 163
 QY 255 AELLTPSAFGSTTGPRA---HMEVLLRARALETQCYVVAAGCGRHHEKASGHSWVD 311
 DB 164 AELLTPSAFW--VRGPKHEHMTSLLAARALDITTCYVVAAGCGC---NKNIGOSRTID 216
 QY 312 PMGTVAACSEGGPLCLARIDLNYLRQLRHLRPFQHR 350
 DB 217 PFGVTIAASEMPALLMAEVTPEPVQVRAQLVPLNRR 255

RESULT 7
 Y480_MYCTU
 ID Y480_MYCTU STANDARD: PRT; 340 AA.
 AC 011146;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 35.2 KDA PROTEIN YV0480C.
 GN YV0480C OR M70498 OR M7C2069.06C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_Taxid=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA MEDLINE=98293987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLaren J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RP SEQUENCE FROM N.A.

RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., Deboy R., Dodson R., Gwim M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Biswal W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 277162; CAB00941.1; -
 DR EMBL: AE006951; AAK44721.1; -
 DR TIGR: MT0498; -
 DR TubercuList; RV0480c; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 DR Hypothetical protein; Complete proteome.
 KW DOMAIN
 FT 192 201 VAL-RICH.
 SQ SEQUENCE 340 AA; 35195 MW; 541EAC9CD1D439B9 CRC64;

Query Match 13.7%; Score 262.5; DB 1; Length 340;
 Best Local Similarity 28.4%; Pred. No. 1.5e-15;
 Matches 89; Conservative 41; Mismatches 140; Indels 43; Gaps 10;

QY 53 PGLRIQLSLVCAOPRPRAAIISSSCSELPVAVCOVTPRDKQONFKTCAELVEARL 112
 DB 42 PGSCARCCALVAGPRLARHRI-----ALAQIRSTIDPAANIQLVYKTAGENATA 91
 QY 113 GAGIAPLPEAFDFIARDPAETLHLSPLGKLEETOLARECGMLSLGCFHERGODWE 172
 DB 92 GAGLVVEPER--TMCRLGVYLRQVAEPDGMANGVRIATEGIVIAEMFFPTGGG-- 147
 QY 173 QIQKINCHVLL--NSKGAIVATYRKTHLQDVETPGGPMCESNSTMPGSLSPVSTPA 230
 DB 148 ---RVNTNLTIAAGPTPNQPDAAHYKHILHYDA-----FGFTESRTVAPG--REPVVVV 196
 QY 231 G--KIGLAVCYDMRFPSELALAOAGAEILTYPSANGSITGPA--HWEVLARALAEIQ 285
 DB 197 DGVRVGLTVCYDIRFPALYTELARGAQLTAVCASWS--GPGKLEQWTLIARALDSM 254
 QY 286 CYVVAAGC-----GRHHEKRASYGSHMVVDPMGTVAACSEPGICLARIDLNTL 336
 DB 255 SYVAAAGADPDGARTGVGASSAAPTGVGSLVASPLGEVYVNASGTQPOLLVADIDVNV 314
 QY 337 RQLRHLPVYQHR 349
 DB 315 AAARDRIAVLRNQ 327

RESULT 9
 YPOO_PSEFL STANDARD; PRT; 285 AA.
 ID YPOO_PSEFL
 AC P55176;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOHETICAL 31.2 KDA PROTEIN IN PPOF 5'REGION (ORF2).
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NCBI_TaxID=294;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CHA0.
 RX MEDLINE=96064397; PubMed=8526497;
 RA Schneider U., Keel C., Defago G., Haas D.;
 RT "Tns-directed cloning of ppg genes from Pseudomonas fluorescens CHA0;
 RT mutational inactivation of the genes results in overproduction of the
 RT antibiotic pyoluteorin";
 RL Appl. Environ. Microbiol. 61:3856-3864(1995).
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X87299; CA60729.1; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 285 AA; 31163 MW; 68B7C64F38CBDE8 CRC64;

Query Match 13.2%; Score 254.5; DB 1; Length 285;
 Best Local Similarity 31.5%; Pred. No. 6.1e-15;
 Matches 92; Conservative 43; Mismatches 126; Indels 31; Gaps 13;

QY 74 ISSSCSELPVAVCOVTPRDKQONFKTCAELVEARLGAAGIAPLPEAFDIARDPAE 132
 DB 14 LSVSGVTMTVAALQCPPLDVAAGNLRQHVAMEAT--DADLVLPPEMLSGSYNGLE 71
 QY 133 TL-HLSEPLGKLEETOLARECGMLSLGCFHERGODWEQTKYNCHVLSKGANV 191
 DB 72 AVGALAEADGSPSAQRIATAAQAATAI-LYGYERSYDGO---IYNANVLDAQQR 126
 QY 192 ATYKTHLC-DVEITPGGPMCESNSTMPGSLSP-VSTPAKIGLAVCYDMRFPSELTA 249
 DB 127 CNVKTHTLFGD-----HSMFSAGEDDPVLVLDOSWKIGPLCYIEEPENAR 176
 QY 250 LAQGAELITYPSAFSGITGPAHWEVLIRARAITQCYVVAAGCGRHHKASY-GHSM 308
 DB 177 LALAGAEILIVPTA-NMIPYDFVADVTIRAFENOCYVAVANYCG--HEEOIRYCGOSS 233
 QY 309 VVDPMGTVAACSEPGICLARIDLNTL--RQLRHLPVYQHRPDLXGML 357
 DB 234 TAAPDGRIALAGIDEALITIGTIDRLMGESRALNRYL---SDRRPELYDL 282

RESULT 9
 YAG5_STALU STANDARD; PRT; 234 AA.
 ID YAG5_STALU
 AC P55178;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOHETICAL PROTEIN IN AGR OPERON (ORF 5) (FRAGMENT).
 OS Staphylococcus lugdunensis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 NCBI_TaxID=28035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93366158; PubMed=8359673;
 RA Vandenesch F., Projan S.J., Kreiswirth B., Etienne J., Novick R.P.;
 RT "Agt-related sequences in Staphylococcus lugdunensis";
 RL FEMS Microbiol. Lett. 111:115-122(1993).
 CC -1- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; F13334; AAA17975.1; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR001110; UPF0012.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS01227; UPF0012; 1.
 DR Hypothetical protein.
 FT NON_TER
 KW
 SQ SEQUENCE 234 AA: 26495 MW: C1CBADCA5E1389A7 CRC64;

Query Match 12.0%; Score 231; DB 1; Length 234;
 Best Local Similarity 31.1%; Pred. No. 5,3e-13;
 Matches 61; Conservative 27; Mismatches 64; Indels 44; Gaps 6;

QY 177 IYNCVHLNLSKGAIVATYRTHLCVDYEIFGQGPMSNSTMPGSLSPSTPAG-----231
 DB 64 IFTFAFDKTKGVINQYDKMLV-----PMLEDEPAFLTAGKNVPE 104
 QY 232 -----RIGLAVCYDMRPFELSLAQAAGAILTY---PSAFSITGPAHMEVLT 279
 DB 105 TFKLSNGVAVTQMCYDLDLPPELLRRPARSGATIAFYVAGWPSARLN-----HWQVLLRA 159
 QY 280 RALETQCYVAAAACGRHHEKRASY-GHSMVVDPMCTVVARCSEGFGLCIARDLVLYRO 338
 DB 160 RALENMNVYVIGNGCG--YDGKTQYAGHSAVINPNCETIIQELSTKELTVTIDIDAVEQ 217
 QY 339 LRRLHPVFQHRREDLY 354
 DB 218 ORKALPVPDSLPVPHLY 233

RESULT 10
 NR13_ARATH STANDARD; PRT; 346 AA.
 AC P46010; 004909;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 21-NOV-1995 (Rel. 32, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NITRILASE 3 (EC 3.5.5.1)
 OS NIT3 OR AT3G44320 OR T10D17.110.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=94294436; PubMed=8022831;
 RA Barcel B., Fink G.R.;
 RT "Differential regulation of an auxin-producing nitrilase gene family
 in Arabidopsis thaliana.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=98145459; PubMed=9484465;
 RA Hillebrand H., Bartling D., Weller E.W.;
 RT "Structural analysis of the nit2/nit1/nit3 gene cluster encoding
 nitrilases, enzymes catalyzing the terminal activation step in indole-
 acetic acid biosynthesis in Arabidopsis thaliana.";
 RL Plant Mol. Biol. 36:89-99(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;

MEDLINE=21016720; PubMed=11130713;
 RA Salimounat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,
 RA Farnham B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delany M., Boutry M., Grivell L.A., Maché R., Puidomenech P.,
 RA De Simone V., Choisy N., Artiguenave F., Robert C., Brotier P.,
 RA Winkler P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
 RA Schaefer M., Meiller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,
 RA Weidemann R., Kranz H., Voss H., Holland R., Brandt P.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstrek G.,
 RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwalder B., Duchemin D.,
 RA Cooke R., Laude M., Berger-Llauró C., Purnelle B., Masny D.,
 RA de Haan M., Maarse A.C., Alcazar J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argilou A., Flores M., Liguori R., Vitale D.,
 RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Wails A., Utterback T., Fujii C.T., Shea T.P.,
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Millscher J., Sellers P., Gill J.E., Feldlyum T.V.,
 RA Preuss D., Lin X., Nieman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Frazer D., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana".
 RL Nature 408:820-822(2000).
 CC - FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
 CC INDOLE-3-ACETIC ACID.
 CC - CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
 CC - SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.
 CC - SIMILARITY: BELONGS TO THE NITRILASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U09959; AAA19627.1; -
 DR EMBL; Y07648; CAA68936.2; -
 DR EMBL; AL353865; CAB89000.1; -
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR000132; Nitril_cyn_hydrolase.
 DR Pfam: PF00795; CN_hydrolase; 1.
 DR PROSITE: PS00920; NITRIL_CHT_1; 1.
 DR PROSITE: PS00921; NITRIL_CHT_2; 1.
 KW Hydrolase; Multigene family; BY SIMILARITY.
 FT ACT_SITE 186
 SQ SEQUENCE 346 AA: 38022 MW: 70CFE421547F2B5E CRC64;

Query Match 11.5%; Score 221.5; DB 1; Length 346;
 Best Local Similarity 27.4%; Pred. No. 5.6e-12;
 Matches 86; Conservative 44; Mismatches 113; Indels 71; Gaps 14;

QY 91 STPDKOONFKTCAELVREARALGACLAFLPEAF-----DFTARDPAET 133
 DB 41 ATLDKAEKF-----IVEAASKGAKLVLPPEAFIGGYPRGFRGLAVGVHNEGGRDEFNR 94
 QY 134 LHLSE-PLGCKLLEETQTLAREGGLSLGFGHEKQDMEDQTKYINCHVNLNSKAAVVA 192
 DB 95 YHSAKIVGPEVERLAELAGKNNVHLVMALEKDG-----YTLCTALFFSPQGF 148
 QY 193 TYRKTHLCVDE--IPGQGMCSNSTMPGSLSPSTPAGKIGLAVCDMRPELSLAL 250
 DB 149 KHKVMPSTLERCIMGOG-----DGSTIP-----YVDPIGKIGALICENRNPPLYRTAL 198


```

RL Nature 408:820-822(2000).
RN [5]
RP CHARACTERIZATION.
RC STRAIN=CV, LANDSBERG, ERRECTA;
RA MEDLINE=94286570; PubMed=8016109;
RX Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;
RT "Molecular characterization of two cloned nitrilases from Arabidopsis
RT thaliana: key enzymes in biosynthesis of the plant hormone indole-3-
RT acetic acid."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
CC INDOLE-3-ACETIC ACID.
CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT, BUT AT A
CC VERY LOW LEVEL DURING THE FRUITING STAGE.
CC -1- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X63445; CAA45041.1; -
DR EMBL: U38845; AAR05221.1; -
DR EMBL: Y07648; CAA68935.2; -
DR EMBL: AL353865; CAB86999.1; -
DR PIR: S22398; S22398.
DR InterPro: IPR003010; CN_hydrolase.
DR InterPro: IPR000132; Nitril_cyn_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS00920; NITRIL_CHT_1; 1.
DR PROSITE: PS00921; NITRIL_CHT_2; 1.
DR KW Hydrolase; Multigene family.
FT ACT_SITE 186 186 BY SIMILARITY.
FT CONFLICT 312 312 Y -> H (IN REF. 2).
FT FT 312 312
SQ SEQUENCE 346 AA; 38178 MW; 8D4F8B7CADLE3C1F CRC64;

```

```

AC P32962; Q96505;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NITRILASE 2 (EC 3.5.5.1).
GN NIT2 OR AT3G44300 OR T10D17_90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, LANDSBERG, ERRECTA; TISSUE=Leaf;
RX MEDLINE=94286570; PubMed=8016109;
RA Bartling D., Seedorf M., Schmidt R.C., Weller E.W.;
RT "Molecular characterization of two cloned nitrilases from Arabidopsis
RT thaliana: key enzymes in biosynthesis of the plant hormone indole-3-
RT acetic acid."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6021-6025(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=94294436; PubMed=8022831;
RA Bartel B., Fink G.R.;
RT "Differential regulation of an auxin-producing nitrilase gene family
RT in Arabidopsis thaliana."
RL Proc. Natl. Acad. Sci. U.S.A. 91:6649-6653(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Zhou L., Bartel B., Thornburg R.W.;
RT "Nucleotide sequence of a pathogen induced nitrilase gene from
RT Arabidopsis thaliana. Nit2."
RL (In) Plant Gene Register PGR96-006.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unselid M.,
RA Fartman B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Griveil L.A., Maché R., Puigdomenech P.,
RA De Simone V., Cholsine N., Artiguenave F., Robert C., Brothier P.,
RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurzbach M., Drzonek H., Erle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Clement J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Landie M., Berger-Llauró C., Purnelle B., Masny D.,
RA de Haan M., Maarse A.C., Alcazar J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
RA Manhaupt G., Haase D., Schoof H., Rüd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Watis A., Uteback T., Fujii C.Y., Shea T.P.,
RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltischer J., Sellers P., Gill J.E., Feldlyum T.V.,
RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Ideawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
CC -1- FUNCTION: CAN CONVERT INDOLE-3-ACETONITRILE TO THE PLANT HORMONE
CC INDOLE-3-ACETIC ACID.
CC -1- CATALYTIC ACTIVITY: A NITRILE + H(2)O = A CARBOXYLATE + NH(3).
CC -1- SUBCELLULAR LOCATION: TIGHTLY ASSOCIATED WITH THE PLASMA MEMBRANE.
CC -1- DEVELOPMENTAL STAGE: BARELY DETECTABLE IN YOUNG ROSETTES, BUT IS

```


CC STRONGLY EXPRESSED DURING BOLTING, FLOWERING, AND ESPECIALLY FRUIT
 CC DEVELOPMENT.
 CC -! SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68305; CAA48377.1; -
 DR EMBL: U09958; AAB60275.1; -
 DR EMBL: U38845; AAB05220.1; -
 DR EMBL: AL353865; CAB86998.1; -
 DR PIR: S31969; S31969.
 DR InterPro: IPR003010; CN_hydrolase.
 DR InterPro: IPR00132; Nitril_cyn_hydrolase.
 DR Pfam: PF00795; CN_hydrolase.1.
 DR PROSITE: PS00920; NITRIL_CHT_1; 1.
 DR PROSITE: PS00921; NITRIL_CHT_2; 1.
 KW Hydrolase; Multigene family.
 FT ACT_SITE 179 179 BY SIMILARITY.
 FT CONFLICT 37 37 E -> G (IN REF. 3).
 FT CONFLICT 48 48 S -> T (IN REF. 3).
 SQ SEQUENCE 339 AA: 37153 MW: 06CDE76D2FDC24A7 CRC64:

Query Match 10.3%; Score 198.5; DB 1; Length 339;
 Best Local Similarity 25.9%; Pred. No. 5.5e-10;
 Matches 78; Conservative 47; Mismatches 115; Indels 61; Gaps 12;

QY 94 DKQNFKTCALVREARLACIALFLPEAF-----DPLARDPAETLHL 136
 DB 31 DDPATLEKANKRTIVEMASGSELVPEAFICGPRFGLGVGHNEGRDEFKRYHA 90
 QY 137 SE-PLGKLEEVYQLARCGMLSLGCFHERGQDWEQOKIYNCHVNLNSKAAVATYR 195
 DB 91 SAIKVPGPEVEKLAELGANNVYLVMGALKEKD-----YTLCTALFFSPQGLFKHR 144
 QY 196 KTHLCDVE--IPGQPMCSNSTMPSPLESPTSPACKIGLAVCYDMRPPELSLAQA 253
 DB 145 KALMPTSLERCINGOG-----DGSITP-----YDPTIGKLGAIICENRMLPYRTALYAK 194
 QY 254 GAEIITYPASGISTGPAHWEVLLARAIETOCYVVAAG-CGR-----HH 298
 DB 195 GIELCAPFADGS---KEMQSMHIAIEGCGFVLACQFLRKDFPPHPOVLTFTDWD 250
 QY 299 EKR---ASYGSHMVVDPMGVAVARCS-EGPGICLARIDLNYLRRL-PIFGHRRPD 352
 DB 251 DKEPSIYSGGSVLIISPLGQVLAPNFESEGLITADLDGDVAAKLYDFSVGHYSRPD 310
 QY 353 L 353
 DB 311 V 311

RESULT 14
 ALAM_PSEAE STANDARD: PRT: 346 AA.
 AC P11436:
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALIPHATIC AMIDASE (EC 3.5.1.4).
 GN AMTE OR PA3365.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE.

RC STRAIN=PAC142;
 RX MEDLINE=87219101; PubMed=3108029;
 RA Ambler R.P., Aufferet A.D., Clarke P.H.;
 RT "The amino acid sequence of the aliphatic amidase from Pseudomonas
 RT aeruginosa.";
 RL FEBS Lett. 215:285-290(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87219102; PubMed=3108030;
 RA Brammar W.J., Charles I.G., Matfield M., Liu C.-P., Drew R.E.,
 RA Clarke P.H.;
 RT "The nucleotide sequence of the amE gene of Pseudomonas aeruginosa.";
 RL FEBS Lett. 215:291-294(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Keizer J., Sailer M.H., Hancock R.E.M., Iori S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -! FUNCTION: ALLOWS THE ORGANISM TO GROW ON SHORT CHAIN-LENGTH
 CC ALIPHATIC AMIDES.
 CC -! FUNCTION: ENABLES THE ORGANISM TO USE ACETAMIDE AS BOTH CARBON
 CC AND A NITROGEN SOURCE.
 CC -! CATALYTIC ACTIVITY: A MONOCARBOXYLIC ACID AMIDE + H(2)O = A
 CC MONOCARBOXYLATE + NH(3).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M27612; AAA25697.1; -
 DR EMBL: AE004759; AAG06754.1; -
 DR PIR: A26741; A26741.
 DR InterPro: IPR003010; CN_hydrolase.
 DR Pfam: PF00795; CN_hydrolase.1.
 KW Hydrolase; Complete proteome.
 FT CONFLICT 40 40 E -> D (IN REF. 1 AND 2).
 FT CONFLICT 169 169 G -> P (IN REF. 2).
 FT CONFLICT 317 317 E -> D (IN REF. 1 AND 2).
 SQ SEQUENCE 346 AA: 38495-MW: E19CEB474EB92B93 CRC64;

Query Match 10.2%; Score 195; DB 1; Length 346;
 Best Local Similarity 25.2%; Pred. No. 1.1e-09;
 Matches 76; Conservative 43; Mismatches 129; Indels 54; Gaps 11;

QY 76 SSSCEPLVAVCVSTSPDKQNFKTCALVREARL-----GACIALFEE-A 122
 DB 7 SSSNDYGVAVVNV-----KMPRLHTAAEVLIDNARKIAEMIVGKQGLPGMDLVFPEYS 61
 QY 123 FDIARDPAETLHLSEPLGKLEEVYQLARCGMLSLGCFHERGQDWEQ--TKIYNC 180
 DB 62 LOGIMDPAEMETAAVAIPBEETEIFSRAOKRAVW--GVFSLTGRHHEHPKRAYNT 118
 QY 181 HVLNSGAVVAVYRKTHLCDVEIPGQPMCSNSTMPSPLESPTSPACKIGLAVCYD 240
 DB 119 LVLDNNGEIVOKYRKTI-----PWCPIEGMTIPGQOTYVSEPKMKKISLIICDD 168
 QY 241 MRPELSLALAOAGAEILT-----YPSAFSGISTGPAHWEVLLARAIETOCYVVAAGC 294
 DB 169 GNPPEIWRDCAMKGAELIVRCGQYMYPAKDOY-----MMAKAMAMANNVCVAVANAA 221

THIS PAGE BLANK (USPTO)

A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.
A:Reference number: A85001; MUID:20083488
A:Accession: D85088
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <STO>
A:Cross-references: GB:NC_001268; NID:g7267521; PIDN:CAB78004.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g08790
A:Map position: 4

Query Match 33.6%; Score 645.5; DB 2; Length 316;
Best Local Similarity 44.2%; Pred. No. 1e-48;
Matches 129; Conservative 53; Mismatches 91; Indels 19; Gaps 5;
QY 68 RPRMAISSSCEPL-----VAVCVSTPDKOONFKCALVREAAALGACL 116
DB 12 KPSLETRITSSQIPLMATVNTKVRVAAAMTSMNLTMTFATCSRLOYEAAALAGKL 71
QY 117 AFLPEAFDPIARDAETLHLESPGLKLEEVYOLARECGMLSLGFGHERGQDWEQTOK 176
DB 72 ICFENEFVQDKESEKIAEPIDGPMERYCSLARNSIMLSLGGQERDD-----TH 127
QY 177 IYNCHVLLSKGAVVATYRKTHLCDVEIPGQPMCESNSTMPGSPLESVPSTPAKIGILA 236
DB 128 LCNTHVVIDDAGMIRDTYQKMHLEFDVPGSSYKESSTVPGTKIVS-VDSVPGRLGILT 186
QY 237 VCYDMRPELELALA-QNGAEILTYPSAFSGITGPAHMEVLLRAAETQCYVAAACG 295
DB 187 VCYDLRFKITYOQLREFEKAQVLLVPSAFVYTGSAHWEILLRAALETQCVVIAAAG 246
QY 296 RHHEKRSYSGHVVDPGVTVAARCE--GPGLCLARIDLNTYLRHRLPY 345
DB 247 KHNEKREYGTLLIIDPGVTGRLPRVSTGIYVADIDFSLIDSVRIKMP 298

RESULT 3
T40601
putative nitrlase homolog - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T40601
R:Wood, V.; Rajandream, M.A.; Bartell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, February 1999
A:Reference number: 221940
A:Accession: T40601
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-276 <WGO>
A:Cross-references: EMBL:AL035570; PIDN:CAB37598.1; GSPDB:GN00067; SPDB:SPBC651.02
A:Experimental source: strain 972h; cosmid c651
C:Genetics:
A:Gene: SPDB:SPBC651.02
A:Map position: 2
A:Introns: 49/3; 81/3; 226/3; 232/1
C:Superfamily: hypothetical protein YLR351c

Query Match 30.0%; Score 575.5; DB 2; Length 276;
Best Local Similarity 44.2%; Pred. No. 1e-42;
Matches 126; Conservative 50; Mismatches 86; Indels 23; Gaps 8;
QY 81 LPLVACVSTPDKOONFKCALVREARLGLAFLPEAFDPIARDAETLHLESP 139
DB 1 MTLAAVADLNSSGTLKLAICKELISQAAGKACIFPEPSDITAINSDAILELTHNP 60
QY 140 LGGKLEEVYOLARECGMLSLGFGHERGQDWEQTOKLYNCHVLLNSK-----GAVVA 192
DB 61 DCSKFIKRVRSATKHSIFVNI-CVHE-----PSKYN--KLINSLFIEPLHGLIIS 110
QY 193 TYRTHLCLDVEIPGQPMCESNSTMPGSPLESVPSTPAKIGLAICYMRPELELALA 252

DB 111 RYKSAHLEFVEIKNGPPLKESNTLRCGAILPCKTPLGKVSATCEDIRPEQAIIKRN 170
QY 253 AGAETITPSAFSGITGPAHMEVLLRAAETQCYVAAACGRHHEKRSYSGHVVDP 312
DB 171 MGAHITTPSAFTEKTGAHMEVLLRAAALDSCQYVIAPOGKHNKRSYSGHSMYDP 230
QY 313 WGTVAARCS--GP-GICLARIDLNTYLRHRLPVFQHRPPDY 354
DB 231 WGTVAAGYSDISPMGLIFADLDNLVDHVRITYPLD--RRNDLY 273

RESULT 4
B69109
N-carbamoyl-D-amino acid amidohydrolase - *Methanobacterium thermoautotrophicum* (strain C:Species: *Methanobacterium thermoautotrophicum*
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: B69109
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, K.I.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: B69109
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-272 <MTH>
A:Cross-references: GB:AE000934; GB:AE000666; NID:32622924; PIDN:AB86277.1; PID:3262
C:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1811
C:Superfamily: hypothetical protein YLR351c

Query Match 28.0%; Score 537.5; DB 2; Length 272;
Best Local Similarity 38.4%; Pred. No. 2.3e-39;
Matches 109; Conservative 49; Mismatches 95; Indels 31; Gaps 4;
QY 84 VAVCVSTPDKOONFKCALVREARLGLAFLPEAFDPIARDAETLHLESPGLGK 143
DB 3 LGICOMVTEDEKERNVGNVGAEMVYEAARGARLYLPEMF-----TCTPYVN 49
QY 144 LLEXYTQ-----LARECGMLSLGFGHERGQDWEQTOKITYNCHVLLNSKGA 191
DB 50 LFEYAEEDENGYSIKTRMSIARELGIHLVAGSIEP-----ETPEGITVNTSVMDGNNII 104
QY 192 ATYRTHLCDVEIPGQPMCESNSTMPGSPLESVPSTPAKIGLAICYMRPELELALA 251
DB 105 GKRRVHLFDLINVPEISFRESDSLADSV-TVETPQCYAGVIGICTDMRFPPELSRMA 163
QY 252 QNGAEILTYPSAFSGITGPAHMEVLLRAAETQCYVAAACGRHHEKRSYSGHVVDP 311
DB 164 LGGAELVLFPGAFNMNTTTPAHMRLVRSRALDNCQCYVAAAPARNPSASYAAYGSMIAD 223
QY 312 PWGTVVARSCEGPGICLARIDLNTYLRHRLPVFQHRPPDY 355
DB 224 PWGTVVARSCEGPGICLARIDLNTYLRHRLPVFQHRPPDY 267

RESULT 5
S56907
hypothetical protein YJ126w - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein J0706
C:Species: *Saccharomyces cerevisiae*
C>Date: 05-May-1995 #sequence_revision 08-Sep-1995 #text_change 29-Sep-1999
C:Accession: S56907
R:Czepluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56891
A:Accession: S56907
A:Molecule type: DNA
A:Residues: 1-307 <CZ1>
A:Cross-references: EMBL:Z49401; NID:g1008323; PIDN:CAA89421.1; PID:g1008324; MIPS:YJ

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:54:36 ; Search time 56.94 Seconds
(without alignments) 470.926 Million cell updates/sec

Title: US-09-357-675C-25

Perfect score: 1921

Sequence: 1 PLAAAXLAPRPPDRLRMV.....LPVFQHRPDLVGNLGHPLS 362

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1919	99.9	464	21	AAV68739
2	922	48.0	224	21	AAV68739
3	821	42.7	153	20	AAV10877
4	485.5	25.3	276	22	AAV80984
5	310	16.1	159	21	AAV42967
6	283.5	14.8	266	22	AAV92814
7	270	14.1	299	21	AAV09683
8	256.5	13.4	299	21	AAV49945
9	256.5	13.4	252	22	AAV79700
10	256.5	13.4	252	22	AAV80193
11	241	12.5	220	21	AAV09684

12	241	12.5	220	21	AAV49946	Arabidopsis thalia
13	218	11.3	264	22	AAV81841	S. epidemidis ope
14	217.5	11.3	340	21	AAV44266	Arabidopsis thalia
15	217.5	11.3	346	21	AAV44265	Arabidopsis thalia
16	217.5	11.3	358	21	AAV44264	Arabidopsis thalia
17	199.5	10.4	303	15	AAV46260	Improved Heat-stab
18	199.5	10.4	303	15	AAV46258	Improved Heat-stab
19	199.5	10.4	303	15	AAV46251	Heat-stable carbam
20	199.5	10.4	352	21	AAV05762	Arabidopsis thalia
21	198.5	10.3	303	15	AAV46241	Heat-stable carbam
22	198.5	10.3	303	15	AAV46268	Improved Heat-stab
23	198.5	10.3	339	21	AAV05763	Arabidopsis thalia
24	197.5	10.3	303	15	AAV46269	Improved Heat-stab
25	197.5	10.3	303	15	AAV46270	Improved Heat-stab
26	196.5	10.2	303	15	AAV46249	Heat-stable carbam
27	196.5	10.2	303	15	AAV46257	Improved Heat-stab
28	196.5	10.2	303	15	AAV46255	Improved Heat-stab
29	196.5	10.2	303	15	AAV46261	Improved Heat-stab
30	196.5	10.2	303	15	AAV46259	Improved Heat-stab
31	196.5	10.2	303	15	AAV46243	Heat-stable carbam
32	196.5	10.2	303	15	AAV46244	Heat-stable carbam
33	196.5	10.2	303	15	AAV46265	Improved Heat-stab
34	196.5	10.2	303	15	AAV46267	Improved Heat-stab
35	196.5	10.2	303	15	AAV46272	Improved Heat-stab
36	196.5	10.2	303	22	AAV74625	Escherichia coli d
37	195.5	10.2	303	15	AAV46247	Heat-stable carbam
38	195.5	10.2	303	15	AAV46248	Heat-stable carbam
39	195.5	10.2	303	15	AAV46252	Heat-stable carbam
40	195.5	10.2	303	15	AAV46253	Heat-stable carbam
41	195.5	10.2	303	15	AAV46250	Improved Heat-stab
42	195.5	10.2	303	15	AAV46262	Improved Heat-stab
43	195.5	10.2	303	15	AAV46263	Improved Heat-stab
44	195.5	10.2	303	15	AAV46239	Heat-stable carbam
45	195.5	10.2	303	15	AAV46240	Heat-stable carbam

ALIGNMENTS

RESULT 1	
AAV68739	
ID	AAV68739 standard; Protein; 464 AA.
XX	
AC	AAV68739;
XX	
DT	05-MAY-2000 (first entry)
XX	
DE	Amino acid sequence of the N11 protein.
XX	
KW	N11 gene; nitrlase; tumour suppressor gene; FHIT; chromosome 3p14.2;
KW	FRA3B; cancer; genome allele inactivation.
XX	
OS	Homo sapiens.
OS	Mus sp.
OS	Drosophila melanogaster.
OS	Caenorhabditis elegans.
XX	
FT	Key
FT	Misc-difference
FT	Location/Qualifiers
XX	/note="unspecified amino acid encoded by TNT"
PN	WO200003685-A2.
XX	
PD	27-JAN-2000.
XX	
PF	20-JUL-1999; 99WO-US16366.
XX	
PR	20-JUL-1998; 98US-0093350.
XX	
PA	(UYJE-) UNITV JEFFERSON THOMAS.
XX	
PI	Croce CM;
XX	

DR WPI: 2000-171195/15.
DR N-PSDB: AAZ46101.

PT Novel nitrlase homologs used as diagnostic and therapeutic reagents
PT for the detection and treatment of cancer

XX PS Disclosure: Fig 6; 25pp; English.

CC The present sequence is encoded by the coding region of human, murine,
CC Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
CC human and mouse NIT1 genes are members of an uncharacterised
CC mammalian gene family with homology to bacterial and plant nitrlases.
CC The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
CC for fusion proteins in which the Phit domain is fused with a Nit domain.
CC In mouse and humans, FHIT and NIT are encoded by two different genes,
CC localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
CC human FHIT gene at chromosome 3p14.2, spanning the constitutive
CC chromosomal fragile site FRA3B, is often altered in most common forms
CC of human cancer. The NIT1 protein overcomes the mutated inactivation
CC of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
CC and analogues of them, and antibodies are used as diagnostic and
CC therapeutic reagents for the detection and treatment of cancers.

XX Sequence 464 AA:

Query Match 99.9%; Score 1919; DB 21; Length 464;
Best Local Similarity 100.0%; Pred. No. 3,1e-189;

Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLAAAXLAPDRPPDRTLRLMVLAISSCRYSLSRRRLGFTTRPPHRLSLCPGLRIPL 60

Db 1 plaaaxlapdrppdrtllrmvlaisscrvlsarrprlglftrpphrfslslcpqlrpl 60

QY 61 SVLCAQPRPRMAISSSCCEPLVAVCOVTSFDPKQONFKCAELVREARAGACLAFLP 120

Db 61 svlcqprprmaisssscelplvavcvtsfdpkqnfkcaelvrearagacflap 120

QY 121 EAFDIADPAETLHLSPILGKLEETQTLARECGMLSLGFEHRCODMEQOTKIYNC 180

Db 121 eafdiadpaetlhlspilgkllleeytqlarecglwlsigfhergqdwqtkiync 180

QY 181 HVLNSKCAVAVATYRKTHLCDEVEIPGQPMCESTMPQPSLESFVSTPAKIGLAVCYD 240

Db 181 hvlnskgavavatyrlkthlcdeveipgqpmcesntmpqpslesvstpaqiglavcyd 240

QY 241 MRPEPLSLALNAGAEITLTPSAFSGITGPAHWEVLRAARAEETOCYVAAACGRHNEK 300

Db 241 mrpelpsllalnagaeilttysafigstgpahevllraaraleetocyvvaacqgrhnek 300

QY 301 RASYGSHVAVDPWGTGVAVARCSGPGICLARIDLNLRLPVEQHRPDLVGNLGHF 360

Db 301 rasyghshvavdpwgtgvavarcsegpgiclaridlnlrlpvefqrtrpdlvgnlghp 360

QY 361 LS 362

Db 361 ls 362

RESULT 2
AAB57054 standard; Protein: 224 AA.

XX ID AAB57054 standard; Protein: 224 AA.

XX AC AAB57054;

XX DT 13-MAR-2001 (first entry)

XX DE Human prostate cancer antigen protein sequence SEQ ID NO:1632.

XX KW Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;

XX KW antiviral; gastrointestinal; nephrotropic; antiinfective; gynaecological;
antibacterial; gene therapy; neural; immune; reproductive; renal;

KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.

XX OS Homo sapiens.

XX PN WO20005174-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05988.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI (ROSE/) ROSEN C A.

XX DR Rosen CA, Ruben SM;

XX DR WPI: 2000-587513/55.

XX DR N-PSDB: AAF16257.

PT prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer

XX PS Claim 11; Page 2090-2091; 2338pp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB5363 to AAB57302.

XX The prostate cancer antigens can have neuroprotective, cytostatic,
XX cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
XX nephrotropic, antiinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,

XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.

XX Sequence 224 AA:

Query Match 48.0%; Score 922; DB 21; Length 224;
Best Local Similarity 94.7%; Pred. No. 8e-87;

Matches 178; Conservative 3; Mismatches 3; Indels 4; Gaps 3;

QY 37 LGFTTRPPHRLSLCPGLRIPLSVLCAOPPRAMAISSSCCEPLVAVCOVTSFDPKQ 96

Db 27 lgfttrpphrfslslcpqlrplslvcaqprprmaisssscelplvavcvtsfdpkq 86

QY 97 QNFKTCALVREARAGACLAFLPAEPDFIARDPAETLHLSPILGKLEETQTLARECG 156

Db 87 qnfktcaelvrearagacflpaeidfiairdpaetlhlspilgkllleeytqlarecg 146

QY 157 LMLSLGFEHRCODMEQOTKIYNCVHLNSKCAVAVATYRKTHLCDEVEIPGQPMCESTMP 216

Db 147 lmlslgfehergqdwqtkiyncvhlnskavavatyrlkthlcdeveipgqg-ic-vkat 204

QY 217 MP--GPSL 222

Db 205 lpc1gplv1 212

RESULT 3

XX ID AAY10877 standard; Protein: 153 AA.

XX AC AAY10877;

XX DT 14-MAY-1999 (first entry)

XX Amino acid sequence of a human secreted protein.
 DE
 XX
 KW Secreted protein; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; foetal deficiency; blood disorder;
 KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
 KW renal disease; diabetes; inflammation; allergy; ischemic shock;
 KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
 KW prostate disease; asthma; osteoporosis; arthritis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..46
 FT /note="signal peptide"
 FT Protein 47..153
 FT /note="secreted protein"
 XX
 PN W09907891-A1.
 XX
 PD 18-FEB-1999.
 XX
 PF 04-AUG-1998; 98WO-US16235.
 XX
 PR 19-AUG-1997; 97US-0056732.
 PR 05-AUG-1997; 97US-0054798.
 PR 05-AUG-1997; 97US-0054803.
 PR 05-AUG-1997; 97US-0054804.
 PR 05-AUG-1997; 97US-0054806.
 PR 05-AUG-1997; 97US-0054807.
 PR 05-AUG-1997; 97US-0054808.
 PR 05-AUG-1997; 97US-0054809.
 PR 05-AUG-1997; 97US-0055309.
 PR 05-AUG-1997; 97US-0055310.
 PR 05-AUG-1997; 97US-0055312.
 PR 05-AUG-1997; 97US-0055386.
 PR 05-AUG-1997; 97US-0055311.
 PR 18-AUG-1997; 97US-0055970.
 PR 18-AUG-1997; 97US-0055986.
 PR 19-AUG-1997; 97US-0056365.
 PR 19-AUG-1997; 97US-0056366.
 PR 19-AUG-1997; 97US-0056557.
 PR 19-AUG-1997; 97US-0056370.
 PR 19-AUG-1997; 97US-0056371.
 PR 19-AUG-1997; 97US-0056563.
 PR 19-AUG-1997; 97US-0056731.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Ehner R, Ferrie AM, Greene JM, Janat F, N1 J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
 XX
 DR WPI: 1999-167452/14.
 DR N-PSDB: AAX30398.
 XX
 PT New isolated human genes encoding secreted polypeptides - useful for
 PT diagnosis and treatment of pathological diseases
 XX
 PS Claim 3; Page 304; 331pp; English.
 XX
 CC The specification describes secreted proteins and their corresponding
 CC polynucleotides which are useful for preventing, treating or ameliorating
 CC medical conditions, e.g. by protein or gene therapy. Pathological
 CC conditions can also be diagnosed by determining the amount of the
 CC secreted polypeptides in a sample or by determining the presence of the
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the products, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,

CC prostate diseases, asthma, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
 CC thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 XX
 SQ Sequence 153 AA;

Query Match 42.7%; Score 821; DB 20; Length 153;
 Best Local Similarity 100.0%; Pred. No. 1.2e-76;
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 MCESNSTMPGSLSPSTPAGKIGLAVCYDMRPETSLAOGAEILTPSFGSITG 269
 Db 1 mcesnstmpgslspstpapgkiglavcydmrpetslaogaeiltpsfsgsitg 60
 QY 270 PAHWEVLLRRARLETQCYVAAACGRHBEKRASYSCHSWVDPWGVVAVCSGPGICLA 329
 Db 61 pahwevllrrarletqcyvaaaqgrhbekrasyshsmvvdpwgtvavarcsegpgicla 120
 QY 330 RIDLNTYLRQLRRHLPPVQHRRPDLXGNLGHPLS 362
 Db 121 ridlntylrqlrrhlppvqhrppdlxgnlghpls 153

RESULT 4

AAB80984
 ID AAB80984 standard; Protein; 276 AA.

AC AAB80984;

DT 12-JUN-2001 (first entry)

DE Human hnit3-ase.

KW Human: hnit3-ase; Cushings adrenal gland tumour;

KM nitrile hydrolytic enzyme; nitrilase.

OS Homo sapiens.

PN CN1277998-A.

PD 27-DEC-2000.

PF 30-MAY-2000; 2000CN-0116221.

PR 30-MAY-2000; 2000CN-0116221.

PA (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.

PI Xu X, Qian B, Zhang X;

DR WPI: 2001-245678/26.

DR N-PSDB: AAF84214.

PT Human nitrile hydrolytic enzyme protein and its coding sequence -

PS Claim 2; Page 11; 20pp; Chinese.

CC The present invention relates to human hnit3-ase (nitrilase) protein,
 CC which is expressed in Cushings adrenal gland tumour, and its coding
 CC sequence (AAB84214 and AAB80984). The present invention also relates to a
 CC preparation method of said protein and nucleic acid sequence, and a
 CC method of detecting human hnit3-ase nucleic acid sequence and polypeptide
 CC in sample.
 XX
 SQ Sequence 276 AA;

Query Match 25.3%; Score 485.5; DB 22; Length 276;
 Best Local Similarity 36.7%; Pred. No. 1.2e-41;
 Matches 105; Conservative 46; Mismatches 98; Indels 37; Gaps 6;

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 6568; 246pp + Sequence Listing; English.

XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX Sequence 266 AA:

Query Match 14.8%; Score 283.5; DB 22; Length 266;

Best Local Similarity 32.3%; Pred.No. 7.9e-21; Indels 33; Gaps 10;

Matches 90; Conservative 41; Mismatches 115;

QY 84 VAVQVYTPDKQONFKTCALREARARLACIAFLPEAFDFIARPAETLHL---SEPL 140
DB 3 ialqstnsdkmdnfalrrdaekaegarlvlrpea---tsqsfygrldtgaee 58

QY 141 GCKLLEETVQLARECGMLSLGCFH----ERGDMEQOTOKIYNHVLNKAQVATYR 195
DB 59 dgefstavrkldeldvlnvagmflrpadetvgrg---ektstynnmvlslsag-lhgyyn 114

QY 196 KTHLCEVLEIPGCGPMCESTMPGSPLESPVPACKIGIAYCYDMRPEPLSLALQAQA 255
DB 115 khhlyda-----fgyresdtvxpqdel-vvfevdadlkfgyatcydlrfrpeqfxdlarnga 168

QY 256 ELITVPSARFGSITGPA---HMEVLLRRAIETQCYVVAQAOCGRHHEKR-----ASYGH 306
DB 169 qlivvpltsqgd--gpgklqgwevlpraraldscwlvaaogqarlpreeliderkpgpigh 226

QY 307 SNVDPFWGTVAARCSGPGICLARIDLNLKIRRLPLV 345
DB 227 smvtnphgevlasagypemlladldvsgjaklrealpv 265

RESULT 7

AAAG09683
ID AAAG09683 standard; Protein; 299 AA.

AAAG09683;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 7710.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.

KW	Protein identification: signal transduction pathway; metabolic pathway;	PR	23-JUN-1999;	990S-0140354;
KW	Hydratisation assay; genetic mapping; gene expression control; promoter;	PR	24-JUN-1999;	990S-0140693;
KW	termination sequence.	PR	28-JUN-1999;	990S-0140823;
XX		PR	29-JUN-1999;	990S-0140991;
XX	Arabidopsis thaliana.	PR	30-JUN-1999;	990S-0141287;
OS		PR	01-JUL-1999;	990S-0142154;
XX	EP1033405-A2.	PR	02-JUL-1999;	990S-0142055;
XX		PR	06-JUL-1999;	990S-0142390;
XX		PR	08-JUL-1999;	990S-0142803;
PD	06-SEP-2000.	PR	09-JUL-1999;	990S-0142920;
XX		PR	12-JUL-1999;	990S-0142927;
XX		PR	13-JUL-1999;	990S-0143542;
PE	25-FEB-2000; 2000EP-0301439.	PR	14-JUL-1999;	990S-0143624;
PR	25-FEB-1999;	PR	15-JUL-1999;	990S-0144005;
XX		PR	16-JUL-1999;	990S-0144085;
PR	05-MAR-1999;	PR	16-JUL-1999;	990S-0144086;
PR	09-MAR-1999;	PR	19-JUL-1999;	990S-0144325;
PR	23-MAR-1999;	PR	19-JUL-1999;	990S-0144331;
PR	25-MAR-1999;	PR	19-JUL-1999;	990S-0144333;
PR	29-MAR-1999;	PR	19-JUL-1999;	990S-0144334;
PR	01-APR-1999;	PR	19-JUL-1999;	990S-0144335;
PR	06-APR-1999;	PR	20-JUL-1999;	990S-0144332;
PR	08-APR-1999;	PR	20-JUL-1999;	990S-0144384;
PR	16-APR-1999;	PR	21-JUL-1999;	990S-0144814;
PR	19-APR-1999;	PR	21-JUL-1999;	990S-0144881;
PR	21-APR-1999;	PR	21-JUL-1999;	990S-0145086;
PR	23-APR-1999;	PR	22-JUL-1999;	990S-0145085;
PR	28-APR-1999;	PR	22-JUL-1999;	990S-0145087;
PR	30-APR-1999;	PR	22-JUL-1999;	990S-0145089;
PR	04-MAY-1999;	PR	23-JUL-1999;	990S-0145119;
PR	05-MAY-1999;	PR	23-JUL-1999;	990S-0145115;
PR	06-MAY-1999;	PR	23-JUL-1999;	990S-0145114;
PR	07-MAY-1999;	PR	23-JUL-1999;	990S-0145218;
PR	11-MAY-1999;	PR	23-JUL-1999;	990S-0145224;
PR	14-MAY-1999;	PR	26-JUL-1999;	990S-0145276;
PR	14-MAY-1999;	PR	27-JUL-1999;	990S-0145913;
PR	14-MAY-1999;	PR	27-JUL-1999;	990S-0145918;
PR	18-MAY-1999;	PR	27-JUL-1999;	990S-0145919;
PR	19-MAY-1999;	PR	28-JUL-1999;	990S-0145951;
PR	20-MAY-1999;	PR	02-AUG-1999;	990S-0146386;
PR	21-MAY-1999;	PR	02-AUG-1999;	990S-0146388;
PR	24-MAY-1999;	PR	02-AUG-1999;	990S-0146389;
PR	25-MAY-1999;	PR	03-AUG-1999;	990S-0147003;
PR	27-MAY-1999;	PR	03-AUG-1999;	990S-0147204;
PR	28-MAY-1999;	PR	04-AUG-1999;	990S-0147302;
PR	01-JUN-1999;	PR	05-AUG-1999;	990S-0147192;
PR	03-JUN-1999;	PR	05-AUG-1999;	990S-0147260;
PR	04-JUN-1999;	PR	06-AUG-1999;	990S-0147303;
PR	07-JUN-1999;	PR	06-AUG-1999;	990S-0147416;
PR	08-JUN-1999;	PR	09-AUG-1999;	990S-0147493;
PR	10-JUN-1999;	PR	09-AUG-1999;	990S-0147935;
PR	10-JUN-1999;	PR	10-AUG-1999;	990S-0148171;
PR	14-JUN-1999;	PR	11-AUG-1999;	990S-0148319;
PR	16-JUN-1999;	PR	12-AUG-1999;	990S-0148341;
PR	17-JUN-1999;	PR	13-AUG-1999;	990S-0148565;
PR	17-JUN-1999;	PR	13-AUG-1999;	990S-0148684;
PR	18-JUN-1999;	PR	16-AUG-1999;	990S-0149368;
PR	18-JUN-1999;	PR	17-AUG-1999;	990S-0149175;
PR	18-JUN-1999;	PR	18-AUG-1999;	990S-0149426;
PR	18-JUN-1999;	PR	20-AUG-1999;	990S-0149723;
PR	18-JUN-1999;	PR	20-AUG-1999;	990S-0149722;
PR	18-JUN-1999;	PR	20-AUG-1999;	990S-0149929;
PR	18-JUN-1999;	PR	23-AUG-1999;	990S-0149902;
PR	18-JUN-1999;	PR	23-AUG-1999;	990S-0149930;
PR	18-JUN-1999;	PR	25-AUG-1999;	990S-0150566;
PR	18-JUN-1999;	PR	26-AUG-1999;	990S-0150884;
PR	18-JUN-1999;	PR	27-AUG-1999;	990S-0151065;
PR	21-JUN-1999;	PR	27-AUG-1999;	990S-0151086

```
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 14.1%; Score 270; DB 21; Length 299;
Best Local Similarity 27.6%; Pred. No. 2.3e-19;
Matches 85; Conservative 47; Mismatches 124; Indels 52; Gaps 10;

```
OY 83 LVAVQVSTPDKQONPFTCAELVREARLGACLAFLPEAF-----DFTARDPA 131
DB 11 vvsslqfscsdscstnvaaaerlveaahkgnalllqellfegyyfcqagredfkrakp 70
OY 132 ETLHLSEPLGKILREYQOLARECGIMLSLGGFHERGQDMEQOTKRIYNCHVILNSGAVV 191
DB 71 yknh-----ptlarmqklakelgvnvpvstfe-----eantahynsialiddgdcl 117
OY 152 ATYRTHLCDVEIFPGQCPWCSNSTM-FQPSLESVSTPAGTIGLAVCYDMFPPELISLAL 250
DB 118 glytrkshlpd-----gpyggekfyfnpgdtgfkvfqtkfaklgyaalcwdqyffeararam 171
OY 251 AQAAGAEILTYPSAFGS-----ITGPAHMEVLLRARATETQCVYVVAAGOCGR-----HH 298
DB 172 vlqgaeallfypalaigsepqdgldsrldhrrwmqghaganvvlvasnllxkellletelng 231
OY 299 EKRAAS-YGHSNWDVDPGWVAVANCSE-GPGCLARIDLVYLRQLRHLPLVFOHRRPDLX-- 354
DB 232 psqtlfytgstflagptglgtvaeadkksaevlvaqfdlmlkkskrgswyffirtrpdllykv 291
OY 355 -----GNL 357
DB 111
```

```
DB 292 llumdgml 299
RESULT 9
ID AAB79700 standard; Protein; 252 AA.
AC AAB79700;
DT 30-APR-2001 (first entry)
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:134.
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinoic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.
OS Corynebacterium glutamicum.
XX
XX WO200100843-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB00923.
XX
XX 25-JUN-1999; 99US-0141031.
XX 01-JUL-1999; 99DE-1030476.
XX 02-JUL-1999; 99US-0142101.
XX 08-JUL-1999; 99DE-1031415.
XX 08-JUL-1999; 99DE-1031418.
XX 08-JUL-1999; 99DE-1031419.
XX 08-JUL-1999; 99DE-1031420.
XX 08-JUL-1999; 99DE-1031424.
XX 08-JUL-1999; 99DE-1031428.
XX 08-JUL-1999; 99DE-1031434.
XX 08-JUL-1999; 99DE-1031435.
XX 08-JUL-1999; 99DE-1031443.
XX 08-JUL-1999; 99DE-1031453.
XX 08-JUL-1999; 99DE-1031457.
XX 08-JUL-1999; 99DE-1031465.
XX 08-JUL-1999; 99DE-1031478.
XX 08-JUL-1999; 99DE-1031510.
XX 08-JUL-1999; 99DE-1031541.
XX 08-JUL-1999; 99DE-1031573.
XX 08-JUL-1999; 99DE-1031592.
XX 08-JUL-1999; 99DE-1031632.
XX 08-JUL-1999; 99DE-1031634.
XX 08-JUL-1999; 99DE-1031636.
XX 09-JUL-1999; 99DE-1032125.
XX 09-JUL-1999; 99DE-1032126.
XX 09-JUL-1999; 99DE-1032130.
XX 09-JUL-1999; 99DE-1032186.
XX 09-JUL-1999; 99DE-1032206.
XX 09-JUL-1999; 99DE-1032227.
XX 09-JUL-1999; 99DE-1032228.
XX 09-JUL-1999; 99DE-1032229.
XX 09-JUL-1999; 99DE-1032230.
XX 14-JUL-1999; 99DE-1032922.
XX 14-JUL-1999; 99DE-1032926.
XX 14-JUL-1999; 99DE-1032928.
XX 14-JUL-1999; 99DE-1033004.
XX 14-JUL-1999; 99DE-1033005.
XX 14-JUL-1999; 99DE-1033006.
XX 12-AUG-1999; 99US-0148613.
XX 27-AUG-1999; 99DE-1040764.
XX 27-AUG-1999; 99DE-1040765.
XX 27-AUG-1999; 99DE-1040766.
XX 27-AUG-1999; 99DE-1040832.
XX 31-AUG-1999; 99DE-1041378.
XX 31-AUG-1999; 99DE-1041379.
XX 31-AUG-1999; 99DE-1041380.
```


PA (BADI) BASF AG.
XX
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX WPI: 2001-137957/14.
DR N-PSDB: AAF72312.
XX
PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
PS Claim 20; Page 1678-1679; 1737pp; English.
XX
CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ Sequence 252 AA;

Query Match 13.4%; Score 256.5; DB 22; Length 252;
Best Local Similarity 32.5%; Pred. No. 4.5e-18;
Matches 86; Conservative 36; Mismatches 110; Indels 33; Gaps 10;

QY 93 NFKTCAELVREARAGACLAFLPEAFEDFIARDPAETLH---SEPLGKLEFYQLARE 154
Db 3 nfallrdaaekaaegavlvfpea---tsqsfyglrldtgaeldgfeslavvklade 58
QY 155 GGLMLSLGFPH---ERGDMEOTKQIKYNCHVLNSKGAVALVATRKTHLDCVEIRGQCP 209
Db 59 ldvvlvagmfpradvtvqg---ekistvmtvlisgag-lhgyynkhltyda-----fg 109
QY 210 MCESNSTMPGSPLESPTPAKIGILAVCYDMRPPELSLALAQAGAEILTYPSAFGSITG 269
Db 110 yresdtvpgdel-vvfevddkfgvatcgydlrfpeqfkdlarngaqdlivvplswgd--g 166
QY 270 PA---HMEVLBARALEFQCIVVAAAGCGRHNEK-----ASYHSNWDPMGTVAVARC 320
Db 167 pqlleqvewlvpraldstcwlvacggarlpeelrderkkprtyghsmvtnphgevlasa 226
QY 321 SEGPGICLARIDLNYLRDLRHLPV 345
Db 227 gyepemliadivsglakirealpv 251

RESULT 11
AAG09684
ID AAG09684 standard; Protein: 220 AA.
XX
AC AAG09684;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7711.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.

FD	06-SEP-2000.		PR	06-JUL-1999;	9905-0142390.
XX			PR	08-JUL-1999;	9905-0142803.
PF	25-FEB-2000;	2000EP-0301439.	PR	09-JUL-1999;	9905-0142920.
XX			PR	12-JUL-1999;	9905-0142977.
PR	25-FEB-1999;	9905-0121825.	PR	13-JUL-1999;	9905-0143542.
PR	05-MAR-1999;	9905-0123180.	PR	14-JUL-1999;	9905-0143624.
PR	09-MAR-1999;	9905-0123548.	PR	15-JUL-1999;	9905-0144005.
PR	23-MAR-1999;	9905-0125788.	PR	16-JUL-1999;	9905-0144086.
PR	29-MAR-1999;	9905-0126264.	PR	19-JUL-1999;	9905-0144305.
PR	01-APR-1999;	9905-0127462.	PR	19-JUL-1999;	9905-0144331.
PR	06-APR-1999;	9905-0128234.	PR	19-JUL-1999;	9905-0144333.
PR	08-APR-1999;	9905-0128714.	PR	19-JUL-1999;	9905-0144333.
PR	16-APR-1999;	9905-0129845.	PR	19-JUL-1999;	9905-0144334.
PR	19-APR-1999;	9905-0130077.	PR	19-JUL-1999;	9905-0144335.
PR	21-APR-1999;	9905-0130449.	PR	20-JUL-1999;	9905-0144352.
PR	23-APR-1999;	9905-0130510.	PR	20-JUL-1999;	9905-0144632.
PR	23-APR-1999;	9905-0130891.	PR	20-JUL-1999;	9905-0144884.
PR	28-APR-1999;	9905-0131448.	PR	21-JUL-1999;	9905-0144814.
PR	30-APR-1999;	9905-0132048.	PR	21-JUL-1999;	9905-0145084.
PR	04-MAY-1999;	9905-0132407.	PR	21-JUL-1999;	9905-0145088.
PR	04-MAY-1999;	9905-0132484.	PR	22-JUL-1999;	9905-0145085.
PR	05-MAY-1999;	9905-0132485.	PR	22-JUL-1999;	9905-0145087.
PR	06-MAY-1999;	9905-0132486.	PR	22-JUL-1999;	9905-0145089.
PR	06-MAY-1999;	9905-0132487.	PR	22-JUL-1999;	9905-0145192.
PR	07-MAY-1999;	9905-0132863.	PR	23-JUL-1999;	9905-0145145.
PR	11-MAY-1999;	9905-0134256.	PR	23-JUL-1999;	9905-0145145.
PR	14-MAY-1999;	9905-0134218.	PR	23-JUL-1999;	9905-0145224.
PR	14-MAY-1999;	9905-0134219.	PR	26-JUL-1999;	9905-0145226.
PR	14-MAY-1999;	9905-0134421.	PR	27-JUL-1999;	9905-0145913.
PR	14-MAY-1999;	9905-0134370.	PR	27-JUL-1999;	9905-0145918.
PR	18-MAY-1999;	9905-0134768.	PR	27-JUL-1999;	9905-0145919.
PR	19-MAY-1999;	9905-0134941.	PR	28-JUL-1999;	9905-0145951.
PR	20-MAY-1999;	9905-0135124.	PR	02-AUG-1999;	9905-0146386.
PR	21-MAY-1999;	9905-0135353.	PR	02-AUG-1999;	9905-0146388.
PR	24-MAY-1999;	9905-0135629.	PR	02-AUG-1999;	9905-0146389.
PR	25-MAY-1999;	9905-0136021.	PR	03-AUG-1999;	9905-0147038.
PR	27-MAY-1999;	9905-0136392.	PR	04-AUG-1999;	9905-0147204.
PR	28-MAY-1999;	9905-0136782.	PR	04-AUG-1999;	9905-0147302.
PR	01-JUN-1999;	9905-0137222.	PR	05-AUG-1999;	9905-0147260.
PR	03-JUN-1999;	9905-0137528.	PR	05-AUG-1999;	9905-0147303.
PR	04-JUN-1999;	9905-0137502.	PR	06-AUG-1999;	9905-0147416.
PR	07-JUN-1999;	9905-0137724.	PR	06-AUG-1999;	9905-0147493.
PR	08-JUN-1999;	9905-0138094.	PR	09-AUG-1999;	9905-0147493.
PR	10-JUN-1999;	9905-0138540.	PR	09-AUG-1999;	9905-0148171.
PR	10-JUN-1999;	9905-0138847.	PR	11-AUG-1999;	9905-0148319.
PR	14-JUN-1999;	9905-0139119.	PR	12-AUG-1999;	9905-0148341.
PR	16-JUN-1999;	9905-0139452.	PR	13-AUG-1999;	9905-0148565.
PR	16-JUN-1999;	9905-0139453.	PR	13-AUG-1999;	9905-0148684.
PR	17-JUN-1999;	9905-0139492.	PR	16-AUG-1999;	9905-0149368.
PR	18-JUN-1999;	9905-0139454.	PR	16-AUG-1999;	9905-0149175.
PR	18-JUN-1999;	9905-0139455.	PR	18-AUG-1999;	9905-0149426.
PR	18-JUN-1999;	9905-0139456.	PR	20-AUG-1999;	9905-0149722.
PR	18-JUN-1999;	9905-0139457.	PR	20-AUG-1999;	9905-0149723.
PR					


```
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 12.5%; Score 241; DB 21; Length 220;
Best Local Similarity 29.6%; Pred. No. 1.5e-16;
Matches 68; Conservative 41; Mismatches 87; Indels 34; Gaps 8;
```

```
OY 150 QIARECGMLSLGCFHFGMGQEQTKYINCHVLLNSKGAIVATYRKTHLCVEIIPGQCP 209
DB 3 klakelgvivivsvffe-----eantahynsaiadadqtdlgyrkshpdp-----gp 50
OY 210 MCEENSTW-PGPSLESVPSTPAGKIGLAVCYDMRPPELSLALAQAEILTYPSAFGS-- 266
DB 51 gygekfyfnpqdtgfkvqtkfaklqvaiqcwqwfpeaaravlgaaellfypaigsep 110
OY 267 ---ITGPAHWEVLRLARAIETQCVVAAAOGR-----HHEKRAS-YGHSVVVDPMCT 315
DB 111 qdgsldrdhrrvmqngagavnpvlvasnrlgkeltetehpsqtlftgysfllagpge 170
OY 316 VVACSCG-GPELCLARIDNLRRLPVRQHRPDIY-----GNL 357
DB 171 lvaeadkseeavlvagfdlmlkkskrsgwyvfrldirpdllykllmdgnl 220
RESULT 13
ID AAG81841 standard; Protein; 264 AA.
XX AAG81841;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO: 776.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis.
XX
```

```
OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAXO ) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX WPI; 2001-316495/33.
XX DR N-PSDB; AAH52691.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 239; 2188pp; English.
XX
XX AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH5371 to
XX AAH5090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH5091 to
XX AAH5098 represent oligonucleotide sequences and primers which are used
XX in the amplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 264 AA:
```

```
Query Match 11.3%; Score 218; DB 22; Length 264;
Best Local Similarity 31.2%; Pred. No. 4.5e-14;
Matches 58; Conservative 35; Mismatches 69; Indels 24; Gaps 7;
OY 177 IYNCHVLLNSKGAIVATYRKTHLCVEIIPGQGM-----CEENSTWPGPSLESVPSTP 229
DB 94 lyltafaiknkelineykvhlv-----pmlreptflcggn-vvpepfylsdqtl- 143
OY 230 AGKIGLAVCYDMRPPELSLALAQAEILTYPSAFSGSINGPAHWEVLRLARAIETQCVYV 269
DB 144 ---lqilcydlrfrpdllytparkgakielfyvaqws-stldmwlslkxaratendilfv 199
OY 290 AAAOCGRHHEKRASY-GHSMVVDPMCTVVVACSCGEGCLARIDNLRRLPVRQHRPDIY 348
DB 200 acnscg--dggntnyagnavltnpgeilldhlddkegvltlthdvlvdqgfyfipfrn 257
OY 349 RRPDIY 354
DB 258 lkphly 263
RESULT 14
ID AAG44266 standard; Protein; 340 AA.
XX AAG44266;
XX
```

DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55427.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD
XX 06-SEP-2000.
PF
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 03-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

```
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160786.
PR 21-OCT-1999; 99US-0160787.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.3%; Score 217.5; DB 21: Length 340;
Best Local Similarity 26.4%; Pred. No. 7,3e-14;
Matches 87; Conservative 47; Mismatches 128; Indels 67; Gaps 14;

OY 76 SSSCEPLVAVQVSTPPKQNFKCALVREARLACIAPRAAF----- 123
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 16 sstvtvtlvtgsstvyv--dtpatlhaekfiyeaaekgakiylfpeaf199yprgf1 73
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
OY 124 -----DFIARDPAETLHSE-PLGSKLBEVYQOLARECGIMLSIGFFHENGQWEDQTKI 177
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 74 avgvhneegridelfrnyhaaakvprpxverlaelagknvhlvmgalekdg-----ylt 127
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
OY 178 YNCHVLLNSKGAHVATYRKTHLCDFE--IPGQGPCESNSTMPGPLESPVSTPAKIGL 235
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 128 yctalfispqgflgkhrvtmptslerc1wgq-----dgsstlp-----vydtp1gk1ga 177
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
OY 236 AWCYMRPFELSLALQAAGAEILITPSAFGSTTGRAHWEVLARARILEQCYYVAAQ-C 294
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 alcwemrmljyrtalyakjelycapladysl---ewqasmhlhavgcivlsahqfc 233
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
OY 295 GR-----HHEKRAYSGHSMVVDPMGTVVARCS-EGPCLCLARIDLNV 335
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
DB 234 krrfepohpylfnldvtdtkehqdvsggsvlisp1gkvlagpnyeseglvradldlgd 293
OY 336 LKOLKRHLPLVFOH-KRPDLXGNL---GHP 360
   ||: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 294 laraklyfdvghyskppdlf-nltvnehp 321

RESULT 15
AAC44265
ID AAC44265 standard; Protein; 346 AA.
XX
AC AAC44265;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55426.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126784.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
```

```
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142153.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147319.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.3%; Score 217.5; DB 21; Length 346;
Best Local Similarity 26.4%; Pred. No. 7.5e-14;
Matches 87; Conservative 47; Mismatches 128; Indels 67; Gaps 14;

QY 76 SSSCEPLVAVCQVSTPQKQNFKTCALVREARIGACIAFLPEAF----- 123
DB 22 sstvtvtlvqgstvyn---dtpatlhkaekfi veaaskgakiylfpeafigyprgtrfgrl 79
QY 124 ----DTIARDPAETLHUSE-PLGSKLLEETVYQTLARECGIWMISLGGFHGRGDWEOTOKI 177
```

```

Db      80 avyv'hneggrdefrnyhasaalkvpqpxverlaelagknvhlvmgaiekdg-----yfl 133
QY      178 YNCHVLNSKGAVVATYRKTHLCVYE--IPGQGPWCESNSTMPGPSLESVPSTPAKIGL 235
Db      134 yctalfispqgqflgkhrkmpstlerclwggg----dgsstlp-----vydtpigklga 183
QY      236 AVCYDMRPPELSIALAOGAEILTYPSAFGSTGPAHMEVLLRARAIEHQCYVVAAG-C 294
Db      184 aicwenrmplyrtalyakyleycaptadysl-----ewqasmihaveggctvlshqfrc 239
QY      295 GR-----HHEKRASYGHSWVDPWGTVVVARCS-EGPGLCLARIDLNY 335
Db      240 krrefpnhpdyllndivdtkenhdpvsgsgsvlispjgkvlagpnyeseglvrtadidlgd 299
QY      336 LRQLRRHLVFOH-RRPDLYGNL---GHP 360
Db      300 iaraklyfdvavghyskpdif-nltvnehp 327

```

Search completed: April 26, 2002, 16:54:37
 Job time: 235 sec

THIS PAGE BLANK (USPTO)

OM of: US-09-357-675C-21 to: GenEmbl:* out_format : pfs
Date: Apr 29, 2002 11:05 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed_p2n_model -DEV=xlh
-O=/cgn2_1/USPIO.spool/US09357675/rnat_29042002_091757_3179/app_query.fasta.1.1208
-DB=GenEmbl -QMT=fastlap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-EGAPOP=6.000 -EGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09357675.@CGN1_1.10065 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPRXY -WAIT -THREADS=1

Search information block:

Query: US-09-357-675C-21
Query length: 327

Database: GenEmbl:*

Database sequences: 1797656

Search length: 1873333701

Search time (sec): 3374.980000

score list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb.pr:AF069987	+ 1748.00	2415.67	3.4e-126	1385	AF069987 Homo sapiens nitrilas
gb.ro:AF069988	+ 1484.00	2049.81	8.2e-106	1388	AF069988 Mus musculus nitrilas
gb.ro:BC021634	+ 1476.00	2038.53	3.5e-105	1365	BC021634 Mus musculus nitrilas
gb.pr:AF069988	+ 1468.50	2018.45	4.6e-104	1379	AF069988 Homo sapiens nitrilas
gb.hlg:AL354714.4	+ 1468.50	1989.30	1.9e-102	110000	Continuation (5 of 7) of AL3
gb.hlg:AL354714.5	+ 1468.50	1989.30	1.9e-102	110000	Continuation (6 of 7) of AL3
gb.hlg:AL3590651	+ 1468.50	1985.56	3.1e-102	167863	AL3590651 Homo sapiens chrom
gb.pr:AL591806	+ 1468.50	1983.97	3.8e-102	200822	AL591806 Human DNA sequence
gb.hlg:AC084821	+ 1105.50	1480.02	4.5e-74	211771	AC084821 Mus musculus chrom
gb.ro:AF069985	+ 1102.50	1509.97	9.6e-76	4481	AF069985 Mus musculus nitrilas
gb.hlg:AC087229	+ 1083.00	1448.95	2.5e-72	213203	AC087229 Mus musculus chrom
gb.hlg:AC105589	+ 1059.00	1417.03	1.4e-70	178418	AC105589 Rattus norvegicus c
gb.ov:AF284575	+ 968.50	1335.67	4.9e-66	1214	AF284575 Xenopus laevis Nit p
gb.in:AF069989	+ 710.50	935.82	5.4e-46	1521	AF069989 Drosophila melanogast
gb.hlg:AC017755	+ 710.50	936.34	8.6e-44	131853	AC017755 Drosophila melanog
gb.in:AC093121	+ 710.50	934.18	1.1e-43	168250	AC093121 Drosophila melanog
gb.hlg:AC006169	+ 710.50	933.91	1.2e-43	173494	AC006169 Drosophila melanog
gb.in:AC105283	+ 710.50	933.09	1.3e-43	174279	AC105283 Drosophila melanog
gb.in:AE003467	+ 710.50	929.10	2.2e-43	190351	AE003467 Drosophila melanog
gb.in:AF069986	+ 702.00	964.86	2.2e-45	1385	AF069986 Caenorhabditis eleg
gb.pr:AF372904	+ 648.50	892.13	2.5e-41	1172	AF372904 Arabidopsis thaliana
gb.hlg:AC010013	+ 632.50	840.14	2.0e-38	34013	AC010013 Drosophila melanog
gb.ba:AE000934	+ 537.50	715.54	1.7e-31	15124	AE000934 Methanobacterium the
gb.pl:SCYJL126W	+ 519.00	711.54	2.9e-31	1307	249401 S.cerevisiae Chromosome
gb.pl:AF284571	+ 519.00	709.80	3.6e-31	1591	AF284571 Saccharomyces cerevis
gb.pl:SPG651	+ 502.00	661.73	1.7e-28	25350	AC055570 S.pombe chromosome 1
gb.ba:AE004861	+ 500.50	667.48	8.1e-29	10466	AE004861 Pseudomonas aerugin
gb.pr:AF670334	+ 493.00	677.78	2.2e-29	1019	AF670334 Homo sapiens CNA002 m
gb.pl:SCGCD14	+ 493.00	668.77	6.9e-29	2792	254149 S.cerevisiae GCD14 gene
gb.ba:STCSLR0	+ 490.00	632.03	7.7e-27	110908	D64002 Synchocystis sp. PCG
gb.pr:BC020620	+ 487.50	670.54	5.5e-29	965	BC020620 Homo sapiens Nit prot
gb.pr:AF284574	+ 485.50	667.79	7.8e-29	962	AF284574 Homo sapiens Nit prot
gb.ro:AF284573	+ 481.00	658.94	2.4e-28	1292	AF284573 Homo sapiens Nit prot
gb.ba:AB017194	+ 479.50	645.32	1.4e-27	4782	AB017194 Plectonema boryanum c
gb.ro:BC020153	+ 479.00	657.89	2.8e-28	1063	BC020153 Mus musculus, RIKEN c
gb.ba:AP003387	+ 477.50	604.76	2.6e-25	34350	AP003387 Nostoc sp. PCC 7120
gb.ba:AF005512	+ 476.50	644.63	1.5e-27	3180	AF005512 Nostoc commune tRNA-I
gb.sts:G72919	+ 471.00	648.81	8.9e-28	847	G72919 MARC 2849-2850.991933517
gb.ba:AL646071	+ 468.00	596.16	7.6e-25	203050	AL646071 Ralstonia solanaceae

seq_name: gb.pr:AF069987
seq_documentation_block: 1385 bp mRNA linear PRI 23-JUL-1998
LOCUS AF069987
DEFINITION Homo sapiens nitrilas 1 (NIT1) mRNA, complete cds.
ACCESSION AF069987
VERSION AF069987.1 GI:3228665
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1385)
REFERENCE
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rotman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrilas and flit homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
JOURNAL
MEDLINE 98337986
REFERENCE
AUTHORS 2 (bases 1 to 1385)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rotman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St. Philadelphia, PA 19107, USA
LOCATION/Qualifiers
1..1385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
1..1385
/gene="NIT1"
77..1060
/note="NIT1"
/gene="NIT1"
/codon_start=1
/product="nitrilas 1"
/protein_id="AAC39907.1"
/db_xref="GI:3228666"
/translation="MIGFTTRPHRFLSLCGRLRIPOLSVLCAPRRPAMASSSC
ELPIYAVCGVTPRDKOQKFKCALVRAARGLAFAPDPIDAPDPTLLHS
EPLGKLEETVQLARECGMLSLGFFHREGDMDQETOKTYNCHVILNSKAVATIR
KTHLDCVEIPGGGPMCESTNMPGLESVPSTPAKISLAVCYDRLPELSLALQA
GAEIITYPEAFSGITGPAHVEYLRLARALETOCYVVAACGRRHERRASYSHSWVD
PGTIVARSESGGLCLRIDINLYLRQLRRHLRVFQHRRPDIYGNLGHPLS"

FEATURES
source
gene
CDS
BASE COUNT 322 a 382 c 357 g 324 t

alignment_scores:
Quality: 1748.00 Length: 327
Ratio: 5.346 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-357-675C-21 x AF069987

Align seg 1/1 to: AF069987 from: 1 to: 1385

1 MetleuglypHeIleHnraYpProHIsArGpHeLseuSerleuLeucy 17
|||||
77 ATGCTGGGCTTCATCACAGCCCTTCACAGATTCCTGCTTCGTC 126
17 sPrgGlyeunrGtllePrgGlnInseuSerValleuLeuGyslaGlnProAdGp 34


```

149 CCAGAACCATGTC...TCATCACTTCTCTGGAGCTGCCCTGTGGCT 195
51 ValCysGlnValThrSerThrProAspLysGlnGlnAspPheLysPhe 67
196 GTGTGCGAGTATACATCACACCAACAGACAGAACTTAAACATG 245
67 salagLueValaLargLueValaLargLueValaLargLueValaL 84
246 TGTGTAGTGTGTTCAGAGAGCTGCCAGAGCTGCTGCTGCTGCTTC 235
84 eubProGlnaLpheAspPheLalaArgAspProaLagLueLueHis 100
296 TGGCTGAGGATTTGACTTATTTGACAGAAACCTGCCGAGACATTA 345
101 LeuSerGlnProLueGlyGlyLysLueLueGlnLueLueLueLue 117
346 CTGTCCGAACCATGAAATGGGATCTTTGGCCAAATATAGCCAGCT 395
117 aargLueGlyLueLueLueLueLueLueLueLueLueLueLueL 134
396 CAGGGAATGTGGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 445
134 LnaSPTrpGlnGlnThrGlnLysLysLysLysLysLysLysLys 150
446 AAGACTGGGAGGAGATCAGAAATCTACAAATGTCATGCTTTGAA 495
151 SerLysGlyAlaValaValaLalaThrLysLysLysLysLysL 167
496 AGCAAGGATCAGTACGTGCGCAGTACAGAGAACACATCTGTGCG 545
167 LglnLueProGlyLueLueLueLueLueLueLueLueLueLue 184
546 AGAGATCCAGGTCAGGAGGCGGATGAGAGAACCACTATACCAAG 595
184 LpProSerLueGlnLueLueLueLueLueLueLueLueLueL 200
596 GAGGCACTTTGAGCACCTGTCAGACACCGGCTGGCAGTTGGCT 645
201 AlaValCysLysLysLysLysLysLysLysLysLysLysL 217
646 GCATATGTTATGACATGCGGTTCCCTGGAATCTTTGAAATGG 695
217 naLagLueLueLueLueLueLueLueLueLueLueLueLue 234
696 AGCTGGGCGCAAAATCTTACTTATCTTACGCTTTGGATGTT 745
234 LpProLalaLysTrpGlnValLueLueLueLueLueLueLue 250
746 GTCCAGCCCACTGGAGGTGCTGCTGCGGCGCGCATTTGAAT 795
251 CysTrpValValaLalaLalaGlnCysGlyArgHisLysLys 267
796 TGTATATATATAGACAGCGAGTGTGACCGCACCATTAAACAG 845
267 aSerTrpGlyHisSerMetValValaLysProTrpGlyLys 284
846 AATTTATGGCATAGCATGCTGTGACCGCTGGGCGACAGTGTG 895
284 rGcysSerGlnGlyProGlyLueCysLueLalaLargLueL 300
896 GCTGTCGAGGAGGAGGCGCTGCTGCTGCTGCTGCTGCTGCT 945
301 LeuArgGlnLueLargLysLysLysLysLysLysLysLysL 317
946 CTACACACAGATGGCGCAACACCTGCTGTGTCAGACGCGAG 995
317 pleuTrpGlyAsnLueGlyHisProLueLys 327
996 CCTTATGGCAGTCTGGTATCATCTCTCT 1026
seq_name: gp_ro:BC021634
seq_documentation_block:

```

```

LOCUS BC021634 1365 bp mRNA linear ROD 22-JAN-2002
DEFINITION Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA.
ACCESSION BC021634
VERSION BC021634.1 GI:18204912
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 1365)
JOURNAL Strausberg, R.
Direct Submission
Submitted (14-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: cgapbs-remail.nih.gov
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbio.org
Anup Madan, Jessica Fahney, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
FEATURES
source
1..1365
/organism="Mus musculus"
/db_xref="taxon:10090"
/db_xref="locus:13825 IMAGE:4008543"
/clone="MGC:13825 IMAGE:4008543"
/tissue_type="Mammary tumor metastasized to lung.
MMTV-1TR/Mnt1 model. Expression driven by an MMTV-1TR
enhancer."
/clone_lib="NCI_CGAP_Lu30"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
59..1030
/codon_start=1
/product="nitrilase 1"
/protein_id="AAH21634.1"
/db_xref="GI:18204913"
/translat="MLGFTTRPHOLLCTGYRLTRPVLCSTQPRPRTSSSTSWELPL
VAVCVSTSTPNKQENFKTCAELVQEAARIGALATLPAFPTIARNPETHLLSPLN
GDLGQYSQALAREGIMLSIGFHEHGDMDQONKTYNCNHYLNSGVSVAYSRTNL
CDVEIPGGQSVMSNVTYRPGFTLEPVTYRPGKVLATCYOHRPELSIKLAQACAEI
LTVSARSGSVGRHMEVYLRLARATESCCYVLAACCCRHHTTASVGHSMVVDWGT
VVARCSSEPGICLARIDILHFLQDMQONLPRVORHRRPDLXSLGHPLS"
BASE COUNT 368 a 334 c 331 g 332 t
ORIGIN
alignment_scores:
Quality: 1476.00 Length: 327
Ratio: 4.824 Gaps: 2
Percent Similarity: 93.578 Percent Identity: 84.098
alignment_block:
US-09-357-675C-21 x BC021634 ..
Align seq 1/1 to: BC021634 from: 1 to: 1365
1 MetLueGlyPheLleThrArgProProHisArgPheLueSerLueLueCy 17
|||||

```

```

59 ATGCTGGGCTTCATCACCAGGCTCCTCACAA.....CTCCTGTG 99
17 SPProGluLeuArgIleProGluLeuSerValLeuCysAlaGlnProArgP 34
100 TACCGGATACCGATTACTTGAACCCCACTACTTGTACTCAGCCAGCC 149
34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
150 CCAGAACCATGTGC...TCATCACTTCTCGGAGCTGCCCTGGGGCT 196
51 ValGlyGlnValThrSerThrProAspGlyGlnGlnAspPheLysThr 67
197 GTGTGCCAGTAACTACACACCAACCAAGCAAGAACATTAAACATG 246
67 saLaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
247 TCTGTAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTCCCTGGCTT 296
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
297 TCCCTGAGGCAATTTGACTTATTTGACGAACCCCTGCCAGACATTATC 346
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeu 117
347 CTGTCCGAACCACTGAATGGGATCTTTTGGCCCAATATATAGCCAGCT 396
117 aArgGluCysGlyLeuThrPheSerLeuGlyGlyPheHisGluArgGly 134
397 CAGGGAAATGTGAATCTGGCTCTCTTGGGCGGTTTCCAGACGGTGGCC 446
134 InAspTPeGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAs 150
447 AAGACGCGGAGCAATCAGAAATCTACAAATGTGATGCTGCTTTGAAC 496
151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAsp 167
497 ACCAAGGAGGATAGTAGTGGCCAGTACAGAAACACACTCTGCGCATGT 546
167 IsLuleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPro 184
547 AAGATCCAGGTCAGGGCGCGATGAGAAACCACTATACCAACCGC 596
184 LProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
597 GAGGACACTGTGAGCCACCTGTCAAGACACCGGCTGCAAGTGGTCTA 646
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 217
647 GCAATCTGTATGACATGGCGTTCCTGAACTTCTTGAATGGCTCA 696
217 naLaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 234
697 AGCTGGGGAGAAATACTTACTTATCTTACACCTTGTGATCTGTACAG 746
234 LProAlaHisThrPgluValLeuLeuArgAlaArgAlaIleGluThrGln 250
747 GTCACAGCCCACTGGAGGTGCTGCGGGCCGCCCATGAATCTGAC 796
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArg 267
797 TCTCTGTGAATAGCAGACCGCAGTGTGACGCCACCAACAAACAGAC 846
267 aSerTyrGlyHisSerMetValValAspProTyrPgluThrValAlaAla 284
847 AAGTTATGGCCATAGCATGTGTGACCCGTGGGGACAGTGGTGGCC 896
284 rGysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
897 GCTGCTCCGAGGAGGACGAGCTCTGCTGCTGCTGCAATGATCTCCACT 946
301 LeuArgGlnLeuArgHisLeuProValLProGlnHisArgArgProAs 317
947 CTACACAGATGCGCCAAACACCTGCTGTGTTCAGACCCGAGACCTGA 996

```

```

317 pleuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
997 CCTCTATGGCACTGTGGTCACTCACTCTCT 1027

seq_name: gb_pr:AF069984

seq_documentation_block:
LOCUS AF069984 4079 bp DNA linear PRI 23-JUL-1998
DEFINITION Homo sapiens nitrlase homolog 1 (NIT1) gene, alternatively spliced
product, complete cds.
ACCESSION AF069984
VERSION AF069984.1 GI:3242977
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4079)
Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,
Tillib, S., Draganesu, A., Wermuth, P., Rothman, J.H., Huebner, K.,
Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.
Nitrlase and flit homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
2 (bases 1 to 4079)
Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,
Tillib, S., Draganesu, A., Wermuth, P., Rothman, J.H., Huebner, K.,
Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
source
location/Qualifiers
1..4079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3267)
/gene="NIT1"
/note="alternatively spliced"
/product="nitrlase homolog 1"
635..3569
/gene="NIT1"
/gene="NIT1"
635..680
/gene="NIT1"
/note="alternatively spliced"
/number=1
join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3267)
/gene="NIT1"
/note="NIT1; alternatively spliced"
/codon_start=1
/product="nitrlase homolog 1"
/protein_id="FAC3901.1"
/db_xref="GI:3242978"
/translation="MISGTRPPHRRFLSLCPGRIPQLSVLCAOPPRAMATSSSC
ELPLVAQVYTPDPKQNFRTCAELVREARAGACAPLPEAFDFIARQPAETLHS
EPLGKLLERYTDLAREGGLMLSGSFHEGQDPEQIKLYNCHVLNLSGAAVATYR
KTHLQDVEIPGQGPMSNSTMPGSPLESPVSPAGIKGLAVCYDKRFPRLSLALQA
GAELITPSPAFSGSTGPAHNEVLLRABAIETOCYVVAACGGRHNEKASVGHSMVD
PMGTVVARCSESGPGLCLARDLNTLRQLRHHLPFQHRRPDLYGNLHPLS"
681..711
/gene="NIT1"
/note="1A; alternatively spliced"
712..741
/gene="NIT1"
/note="1B; alternatively spliced"
940..1053
/gene="NIT1"
/note="1C; alternatively spliced"
exon
exon
exon

```

```

exon      1288..1383
           /gene="NIT1"
           /number=2
exon      1636..1890
           /gene="NIT1"
           /number=3
exon      2011..2114
           /gene="NIT1"
           /number=4
exon      2321..2454
           /gene="NIT1"
           /number=5
exon      2609..2734
           /gene="NIT1"
           /number=6
exon      3001..3569
           /gene="NIT1"
           /number=7

BASE COUNT   940 a   1082 c   1043 g   1012 t   2 others
ORIGIN

```

```

alignment_scores:
    Quality: 1468.50      Length: 660
    Ratio: 4.505          Gaps: 8
    Percent Similarity: 49.394    Percent Identity: 48.939

```

alignment_block:

US-09-357-675c-21 x AF069984 ..

Align seg 1/1 to: AF069984 from: 1 to: 4079

```

2 LeuGlyPheIleThrArgProPheHisArgPheLeuSerLeuLeuCysPr 18
|||||
1289 CTGGGCTTCATCAACAGCCCTCCACAGATTCCTGCTGCTGCTGCC 1338
18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
|||||
1339 TGGACTCCGGATACCTCAACCTCAGTACTTGTGCTCAGCCAGGTAC 1388
32 .....
1389 ACGTTTGTGTGTCCTCAGTGCCTGCGACCTAGATGCTCAGTTGTTAA 1438
32 .....
1439 ATGGATAGTGGAGACACAGAGTGTCACTATCCACACATTGATTGCT 1488
32 .....
1489 GAGCCCTACTAGCCCTGGTCAACGTCCCTGTAGAGCATGATCAAAAG 1538
32 .....
1539 GAAGTCCAGCTTCCCTGCTCTCCACTTGACACCCCTTAGCATTAATTG 1588
33 .....ArgPr 34
1589 CTTCCTGTCGTATGAATCTGAGAACTCGCCATGCTGTTCAAGAGGCC 1638
34 oATGAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuAlaVal 51
|||||
1639 CAGAGCCATGGCTATCTCTCTCTCTCTCTGCAACTGCCCTGGTGGCTG 1688
51 aLcysGlnValIThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
|||||
1689 TGTGCCAGGTAACTGACGCGACAGACAAGCAAGAACTTTAAACAATGT 1738
68 AlagIleuValArgIleuAlaIleArgLeuGlyAlaCysLeuAlaPheLe 84
|||||
1739 GCTGAGCTGTGTCGAGAGCTGCAGACTGGGTGCCCTGGCTTCCCT 1788
84 uProGlnAlaPheAspPheIleAlaIleArgAspProAlaGlnThrLeuHsL 101

```

```

|||||
1789 GCCGTGAGGCAATTGACTTCATTGACAGGGACCTGACAGAGCGCTACACC 1838
101 euSerGlnProLeuGlyGlyLysLeuLeuGlnGluIleThrGlnLeuAla 117
|||||
1839 TGTCTGAACCACTGGGTGGGAACTTTTGAAGAATACACCACCTTGCC 1888
117 .....
1889 AGGTATCAGGGAATAGCGAGGAGAGGTAGAAATCTTGTTCAGACTGT 1938
117 .....
1939 CCCTGGGTTGCCAGATATGAGGAGTAGAGCCTTGAGAGTCAGTAAGAGT 1988
118 .....ArgLcysGlyLeuThrPleuSerLeuGly 127
|||||
1989 TGTCAGTGTCCTCCCTCCCGGGAATGTGGACTCTGCTGCTGCTGGGT 2038
128 GlyPheHisGlnArgGlyGlnAspIleGlnIleThrGlnLysIleTyrAs 144
|||||
2039 GGTTCATGAGCGTGGCCAGAGCTGGAGCAGACTGAGAAATCTACAA 2088
144 nCysHisValLeuLeuAsnSer..Lys.....
|||||
2089 TTGTACAGTGTGCTGCTGAACAGCAAGGTGAGACTTTATACCTTTAGCC 2138
152 .....
2139 TGCCCTTCCCATGCTCTCTTACTAGATTCCTCCAAATGTTTCTCAAC 2188
152 .....
2189 TCTTATTCTTGAACCAAGGATTAGGGGTGTCCTACTTCAGTTCCCTA 2238
152 .....
2239 GCCATTAACATATCTCCTCTGGGAGAGATAGCAAGGCTTCTAGAACAA 2288
153 .....GlyAlaValAlaIleThr 159
|||||
2289 CCAGCACTGATATTCCTTCTTACTGTAGGGGACAGTACAGGCACTT 2338
159 yTArgLysThrHisLeuCysAspValGlnIleProGlyGlnGlyProMet 175
|||||
2339 ACAGAAAGACACATCTGTGACGTAGAGTCCAGGCGAGGCGCTATG 2388
176 CysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerProValSe 192
|||||
2389 TGTGAAGCAACTCTACCATGCTGGGCCAGCTTGTGAGTCACTGTGAG 2438
192 rThrProAlaGlyLysIleGly..Leu.....
|||||
2439 CACACACAGCAGCAAGGTAGAGATTGTGAAGGATGAGAGGAGGAAACAG 2488
200 .....
2489 GAATACTTGAATCGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGTA 2538
200 .....
2539 ATGGAATATGACTAGATGCTGTGACAAACAGAGCAGGAACACTACTAG 2588
201 .....AlaValCysTyrAspMetArg 207
|||||
2589 TAGCTGTTTTCATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2638
208 PheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr 224
|||||
2639 TTCCCTGAACCTCTCTGCGCATTTGGCTCAACGCTGAGCAGAGATACCTAC 2688
224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisThrPleuVal. 240
|||||

```

```
2689 CTATCCTTCAGCTTTGGATCATACAGGCCAGCCCACTGGAGGTAA 2738
241 .Leu..... 241
2739 GATGATGCCCTTTTAAACATAAGGCCCTTTCTTAACCACTCCTCCC 2788
241 ..... 241
2789 CCTGGGCCCTACAGTTAAATTCCTCCCTTTCCACCTAATGGAAAC 2838
241 ..... 241
2839 TCATTCCCGAGATATTCTCTCATGATAGTTAAATAGTAAATCATTC 2888
241 ..... 241
2889 TAGGCATTTACCAAAATAGTCACAATGGGTAGATTGCTGTATATGCC 2938
241 ..... 241
2939 TCACCTGTCACTTCCCACTATTGCTACATGTAATTAAGTGAACACACA 2988
242 .....LeuArgAlaArgAlaIleGluThrGlnCysTy 252
2989 TCTCATGCCCGAGGTGTGCTGCGGGCCGCTGCTATGCAACCCAGTCTA 3038
252 rValValAlaAlaGlnCysGlyArgHisGluLeuArgAlaSer 269
3039 TGTAGTGGCAGCAGCACAGTGTGACGCCCACTGAGAAAGAGCAGATT 3088
269 rGlnHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
3089 ATGGCCACACATGTGTGTAGACCCCTGGGGAACAGTGTGGCCCGCTGC 3138
286 SerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTrpLeuAr 302
3139 TCTGAGGGCCAGGCTCTGCTGCGCCGGAATAGACCTCAACTATCTGCG 3188
302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuAr 319
3189 ACAGTTGCCGCGACACCTGCTGTTCCAGCACCGCAGGCCCTGACCTCT 3238
319 rGlnAsnLeuGlyHisProLeuSer 327
3239 ATGGCAATCTGGGTCAACCCACTGCT 3264

seq_name: gb_hhg:AL354714_4
seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (5 of 7) of AL354714 from base 400001 (AL354714 Homo sapiens chromosome 1)

alignment_scores:
Quality: 1468.50 Length: 660
Ratio: 4.505 Gaps: 8
Percent Similarity: 49.394 Percent Identity: 48.939

alignment_block:
US-09-357-675c-21 x AL354714_4 ..
Align seg 1/1 to: AL354714_4 from: 1 to: 110000
2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPsr 18
```

2002
(radio)
Nov
is 2000

```
103858 CTGGGCTTCATCACACAGGCTCCTCACAGATCTCTGCTGTGCTC 103907
18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
103908 TGGACTCGGATACCTCACTCACTCACTTGTGTCTGACGCCAGGTAC 103957
32 ..... 32
103958 ACCTTTTGTGTGTCTCACTGCTGCTGACCTTAGATGCTAGTTGTAA 104007
32 ..... 32
104008 ATGATAGTGGAGACACAGAGTGTCACTATCACACATTTGATTTGT 104057
32 ..... 32
104058 GAGCCCTACTAGCCCTGGGTCAAGTGCCTGTAAGACATGATCAAAAG 104107
32 ..... 32
104108 GAATGCAGCTTCTGCTCTCCACTGACCCCTTAGCATTAATTTG 104157
33 .....ArgPr 34
104158 CTTCCTGTGCTATGAATCTGAAATCTGCAATGCTGTTCACAGGCC 104207
34 oArgAlaMetAlaIleSerSerSerCysGlnLeuProLeuValAlav 51
104208 CAGAGCATGGTATCTCCTCTCTCTCGGAACTGCCCCCTGTGGCTG 104257
51 aCysGlnValThrSerThrProAspLysGlnIleAsnPheLysTrpCys 67
104258 TGTGCCAGTAAATCATGACGCCAGACACAGACAACTTAACAATGT 104307
68 AlaGlnLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
104308 GCTGAGCTGTTGAGAGGCTGCGACGCTGGGTGCTGCTGCTGCTTCT 104357
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 101
104358 GCCTGAAGCATTTGACTTCAATTCACGGGACCTTCAGAGACCTTAC 104407
104358 GCTTGAAGCATTTGACTTCAATTCACGGGACCTTCAGAGACCTTAC 104407
101 eSerGlnProLeuGlyGlyLysLeuLeuGlnIleTrpThrGlnLeuAla 117
104408 TGTCTGAACCATGGGTGGGAACTTTTGGAGAATACACCAAGCTTGC 104457
117 ..... 117
104458 AGGTATCAGGAAATAGCGAGGAGAGTGAATCTTTTGGACAGTGT 104507
117 ..... 117
104508 CCTGGGTTGCCAGATATAGAGGTAGAGCTTGAGAAAGTACGAAAGT 104557
118 .....ArgLysGlyLeuTrpLeuSerLeuGly 127
104558 TGTCAAGTCCCTTCCCGCCAGGGAATGTGGACTGCGCTGTGGGT 104607
128 rGlnPheHisGluArgGlyGlnAspTrpGlnGlnThrGlnLysIleTyrs 144
104608 GGTTCATAGAGGTGGCCCAAGACTGGAGACAGACTCAAGAAATCTACAA 104657
144 rCysHisValLeuLeuAsnSer...Lys..... 152
104658 TTGTCAAGTGTCTGTGAACAGCAAAAGTGAAGCTTATTAACCTTTAG 104707
152 ..... 152
104708 TGCCTCTCCCATGCTCTTCTACTAGATTCTCCAGATTTGTTCAAC 104757
152 ..... 152
```

```

104758 TCTTATTCTTGACCCAGAGATTAGGGGTGTGCTCACTCACTGTTCTTA 104807
152 ..... 152
104808 GCCTATTAACATATCTCTCTTGGAGAGATGAACAGGCTTCTAGAAC 104857
153 ..... GLAValaValaAlaThrT 159
104858 CCACACATGATATCTCTCTTCTTCTTACGTAGGGGACAGTGTGCGCACTT 104907
159 YTAAGlySThHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
104908 ACAGAGAACACATCTGTGTGACGTAGAGATTCCAGGCGAGGGCGCTATG 104957
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
104958 TGTGAAGCAACTCTACCATCCCTGGGGCCAGTCTGTGATCAGCTGTGAG 105007
192 rThrProAlaGlyLysIleGlyLeu..... 200
105008 CACACCAGCAGGCAAGGTAGAGATTGTGAAGAGATGAGGAGGAGGACAG 105057
200 ..... 200
105058 GAATACCTTGAACGTGCAGTAGAGATGAAGAACCCCTAAGAGAGGGGGTA 105107
200 ..... 200
105108 ATGGAATATGACTAGATGCTGTGACAAACAGAGCAGAGACTACTAAG 105157
201 ..... AlaValCysTyrAspMetArg 207
105158 TAGCGTGTTTTCATTCACAGATTGCTGTGCTGTGCTGTATGACATGGCG 105207
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
105208 TTCCTGAACTCTCTGCGCATTTGGCTCAGCTGAGCAGAGATTCCTTAC 105257
224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
105258 CTATCCCTTACACTTTTGGATCATTAACAGGCCACCCACCTGGAGGTAA 105307
241 Leu..... 241
105308 GATGATGCTTTTAAACATAAGAGGCTTTCTTAACCTCATCTTCCC 105357
241 ..... 241
105358 CCTTGGCCCTACAGTTAAATTCCTCCCTTCCACCTAATGGGAAAC 105407
241 ..... 241
105408 TCATTCGCCAGATATTCTCTCATGATAGTTAAATAGTAATCATTC 105457
241 ..... 241
105458 TAGCAATTACCAAAATAGTCACAATGGGTAGATTGTTGTAATGTCCC 105507
241 ..... 241
105508 TCACCTGTCACTTCCCACTATTTCGTACATGTACTTAAGTGAACACACA 105557
242 ..... LeuArgAlaArgAlaIleGluThrGlnCysTyr 252
105558 TCTCATGCCAGGTGTGCTGCGGGCCGCTGTATCGAAACCCAGTGTGA 105607
252 rValValaAlaAlaGlnCysGlyArgHisGlnIleLysArgAlaSerT 269
105608 TGTACTGGCAGCAGCAGAGTGTGGAGCCACCATAGAGAGAGCAAGTT 105657
269 YrGlyHisSerMetValValaAspProTrpGlyThrValValaAlaArgCys 285
105658 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGGCCGCTGC 105707

```

```

286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302
105708 TCTGAGGGGCCAGGCGCTTGCTGCTGCCCCGAAATAGACTCACTATCTGG 105757
302 GlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
105758 ACAGTTGGCCGACACCTGCTGCTGTTCAGACCGCAGGCGCTGACCTCT 105807
319 YrGlyAsnLeuGlyHisProLeuSer 327
105808 ATGGCAATCTGGGTGTCACCCACTGTCT 105833

seq_name: gb_hlg.AL354714_5
seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments
Fragment Name      Begin      End
AL354714.0         1          110000
AL354714_1         100001     210000
AL354714_2         200001     310000
AL354714_3         300001     410000
AL354714_4         400001     510000
AL354714_5         500001     610000
AL354714_6         600001     699820
Continuation (6 of 7) of AL354714 from Base 500001 (AL354714 Homo sapiens chromosome

alignment_scores:
Quality: 1468.50      Length: 660
Ratio: 4.505          Gaps: 8
Percent Similarity: 49.394      Percent Identity: 48.939

alignment_block:
US-09-357-675c-21 x AL354714_5 ..
Align seg 1/1 to: AL354714_5 from: 1 to: 110000

2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
|||||
3858 CTGGGCTTCATCACACGAGCCTCTCACAGATTCCTGCCCTGTGTCGTC 3907
18 GlnLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
|||||
3908 TGGACTCGGATACCTCACTCTCAGTCTGTGCTCAGCCAGGTAAC 3957
32 ..... 32
3958 ACGTTTGTGTGTCCTCAGTGCCTGGCAGCTTAGATGCTCACTTGTGTA 4007
32 ..... 32
4008 ATGATAGTGGAGACACAGAGTGTCAACTATCCACACATTTGATTGCT 4057
32 ..... 32
4058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAAGACATATCAAAAG 4107
32 ..... 32
4108 GAAGTCCAGCTTTCCTGCGCTCTCAGCTTGACCCCTTAGCATTAATTTG 4157
33 ..... ArgPr 34
4158 CTTCCTGTGTATGAATCTGAGAAATCTGACCTGACCTATGCTGTACAGGCGC 4207
34 GlnArgIleMetAlaIleSerSerSerCysGlnLeuProLeuValAlaVal 51
|||||
4208 CAGAGCCATGCTATCTCTCTTCTCCTGCGAAGCTGCCCTGTGTGGCTG 4257
51 aLcysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
|||||

```

```

4258 TGTGCCAGTAACTGACGCCGACACAGACAACTTTAAACATGT 4307
68 AlagluLeuValArgLualAalArgLeuGlyAlaCysLeuAlaPheLe 84
|||||
4308 GTTAGCTGTGTCGAGAGGCTGCCAGACTGGTGCTGCTGGCTTTCCT 4357
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHsL 101
|||||
4358 GCCTGAGGCATTTGACTTCATTCACGCGGACCTGCAGAGACCTACACC 4407
101 euSerGluProLeuGlyLysLeuLeuGluIuTrThrGlnLeuAla 117
|||||
4408 TGTGTGAACCACTGGGTGGAACTTTGGAAGATACACCAGCTTGCC 4457
117
4458 AGGTATCAGGGAATAGCGAGGAGAGTAGAATCTTTGTGGACACTGT 4507
117
4508 CCCTGGGTTGCCAGATATGAGGGTAGAGCCTTGAGAACTCAGTGAAGAT 4557
118 ArgGluCysGlyLeuTrpLeuSerLeuGly 127
|||||
4558 TGTACAGTGTCCCTTCCCGGAGGAAATGTGACTGTGGCTGTGGTGGGT 4607
128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144
|||||
4608 GGTTCATCAGCGGTGGCCAAAGACTGGGACAGCTCAGAAATCTTCACA 4657
144 ncYHisValLeuLeuAsnSer..Lys..... 152
|||||
4658 TTGTCAAGCTGTGCTGTAACAGCAAGGTGAGACTTATAAACCTTTAGCC 4707
152
4708 TGCCTCTCCCATGCTCTTACCTAGATTCTCAGAAATTGTTCTCAAC 4757
152
4758 TCTTATTCTCTGACCCAAAGATTTAGGGGTGCTTACTTCACTTCTCTA 4807
152
4808 GCCTATTAACATATCTCCTCTGGAGAGAGTAAGCAAGCCTCTAGAAC 4857
153
4858 CCAGCACTGATATTCCTTCTTACTGTAGGGGCGAGTAGTGGCCACTT 4907
159 YrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
|||||
4908 ACAGAAAGAACATCTGTGTGACGTAGAGATTCAGAGGCGAGGCCCTATG 4957
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
|||||
4958 TGTGAAGCAACTCTACATGCTGCGGCCAGCTCTGTGATGACCTGTGAG 5007
192 rTrpProAlaGlyLysIleGly..Leu..... 200
|||||
5003 CACACCAAGCAGGCAAGTAGAGTTGTGAAGATGAGGAGGGAACAG 5057
200
5058 GAATACTTTGACTGGCAGTAGAGAGATAGAAACCCCTAAGAGAGGGGCTA 5107
200
5108 ATGGAATATGACTAGATGCTGTGACAAACAGAGAGAACTACTAAG 5157
201
5158 TAGCGTGTTTTCATTCACAGATTGTGCTGCTGTGCTATGACATGCGG 5207

```

```

208 PheProGluLeuSerLeuAlaLeuAlaGluAlaGlyAlaGluIleLeuTh 224
|||||
5208 TTCCCTGAACCTCTCTCGCATTTGGCTCAAGCTGGAGCAGAGATACTTAC 5257
224 rTyrrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240
|||||
5258 CTATCTTCAGCTTTGTGATTCATTACAGGCCAGCCACTGGGAGGTAA 5307
241 .Leu..... 241
5308 GATGATGCCCTTTTAAACATTAAGGGCCTTTCTTAACCTCATCTTCCC 5357
241
5358 CTTGGCCCTACAGTTAAATTCCTTCCCTTCCACCTAATGGAAAC 5407
241
5408 TCATTTCCCGAGATATTCTCTCATGAATAGTTAAATCATTTCC 5457
241
5458 TAGGCAATTACCAAAATAGTCACATGGGTAGATGCTGTGAATGCC 5507
241
5508 TCACCTGTCACTTCCCATATTGCTACATGTACTTAAGTGAACACACA 5557
242
5558 TGTCAATGCCAGGTGTTGCTGGGCGCGGTATGCAAAACCAAGTCTA 5607
252 rValValAlaAlaAlaGlnCysGlyArgHisLysGluLysArgAlaSerT 269
|||||
5608 TGTAGTGGCAGCAGCAGCAGTGTGACGCCACCATGAGAAAGAGCAAGT 5657
269 YrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
|||||
5658 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGCCCGCTGC 5707
286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTrpLeuAr 302
|||||
5708 TCTGAGGGGCGAGGCCCTCTGCTGCGGAATAGACCTCAACTATCTGCG 5757
302 gGluLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
|||||
5758 ACAAGTGGCCGACACCTGCTGTCTCCAGCACCGCAGGCTGACCTCT 5807
319 YrGlyAsnLeuGlyHisProLeuSer 327
5808 ATGGCAATCTGGGTCACACCACTGCTCT 5833
seq_name: gb_htg:AL590651
seq_documentation_block:
LOCUS AL590651 167863 bp DNA linear HTG 18-AUG-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL590651
VERSION AL590651.4 GI:13990236
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 167863)
AUTHORS Harrison,E.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On May 7, 2001 this sequence version replaced gi:13751001.

```

```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquere@sanger.ac.uk
----- Project Information
Center project name: BA137A12
----- Summary Statistics
Assembly program: XCAP4: version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 164729 bases at least Q40
Consensus quality: 165486 bases at least Q30
Consensus quality: 166169 bases at least Q20
Insert size: 167063; sum-of-contrigs
Insert size: 171878; 3.4% error; agarose-fp
Quality coverage: 8.28x in Q20 bases; sum-of-contrigs Quality
coverage: 8.11x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 9374 9473: contig of 9373 bp in length
* 9474 29891: contig of 20418 bp in length
* 29892 29991: gap of 100 bp
* 29992 36239: contig of 6248 bp in length
* 36240 36339: gap of 100 bp
* 36340 103784: contig of 67445 bp in length
* 103785 103884: gap of 100 bp
* 103885 114878: contig of 10994 bp in length
* 114879 114978: gap of 100 bp
* 114979 126347: contig of 11369 bp in length
* 126348 126447: gap of 100 bp
* 126448 137764: contig of 11317 bp in length
* 137765 137864: gap of 100 bp
* 137865 163160: contig of 25296 bp in length
* 163161 163260: gap of 100 bp
* 163261 167863: contig of 4603 bp in length.
*
Location/Qualifiers
1. 167863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RP11-137A12"
/clone_lib="RP11-11.1"
1. 9373
/note="assembly-fragment:02360
fragment_chain:1
clone_end:17
vector_side:left"
9474..29891
/note="assembly-fragment:01854
fragment_chain:1"
29992..36239
/note="assembly-fragment:00805
fragment_chain:1"
36340..103784
/note="assembly-fragment:01467
fragment_chain:1"
103885..114878
/note="assembly-fragment:00991
fragment_chain:2"
114979..126347
/note="assembly-fragment:01503
fragment_chain:2"
126448..137764
/note="assembly-fragment:01297

```

```

misc_feature      fragment_chain:2"
137865..163160
/note="assembly-fragment:00117
fragment_chain:2"
misc_feature      163261..167863
/note="assembly-fragment:01267
fragment_chain:2
clone_end:sp6
vector_side:right"
BASE COUNT      43927 a 39935 c 40002 g 43197 t 802 others
ORIGIN
alignment_scores:
Quality: 1468.50      Length: 660
Ratio: 4.505          Gaps: 8
Percent Similarity: 49.394      Percent Identity: 48.939
alignment_block:
US-09-357-675C-21 x AL590651/rev ..
Align seg 1/1 to reverse of: AL590651 from: 1 to: 167863
2 LeuGlyPheIleThrArgProPheHisArgPheLeuSerLeuLeuGlyPro 18
|||||
151392 CTGGGCTTCATCACCAGGCTCCTCCACAGATTCTCGTCCCTTGCTGCC 151343
18 oGlyLeuArgIleProGlnLeuSerValLeuGlyAlaGlnPro..... 32
|||||
151342 TGGACTCGGATACCTCAACTCAGTACTTGTGCTCAGCCAGGTAAC 151293
32 ..... 32
151292 ACGTTTGTGTGTGCTCCAGTGCCTGCGCATTAGATGCTCAGTTGTTAA 151243
32 ..... 32
151242 ATGGATAGTGGAGACACAGGAGTGTCAATCATCCACACATTGATTGTT 151193
32 ..... 32
151192 GAGCCCTACTAGCCCTGGGTCACGTCGCTGTAGACATGATCAAAAG 151143
32 ..... 32
151142 GAAGTCAGCTTCCCTGCTCTCCACTTGCACCCCTTAGATTAAATTGG 151093
33 ..... ArgPr 34
151092 CTTCCTGTGCTATGAATCTGAGAAATCTGCTATGCTGTTCACAGGCC 151043
34 oArgAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuValAlaVal 51
|||||
151042 CAGAGCATGGCTATCTCTCTCTCTCTCTCTGCGAACTGCCCTGGGGCTG 150993
51 aLysGlnValThrSerThrProAspLysGlnGlnAspPheLysThrCys 67
|||||
150992 TGTGCCAGGTAACTGACGCGCAGCAAGCAACAACATTAAACACAGT 150943
68 AlaGlnLeuValArgGlnAlaAlaArgLeuGlnAlaCysLeuAlaPhe 84
|||||
150942 GCTGAGCGTGTTCGAGAGGCTGCCAGACTGGTGGCTGCTGGCTTCTCT 150893
84 uProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThrLeuHisL 101
|||||
150892 GCGTAGAGCATTTGACTTCATTGCACGGGACCTCAGACAGGCTTACACC 150843
101 euSerGlnProLeuGlyGlyLysLeuLeuGlnGlnIleuThrGlnLeuAla 117
|||||
150842 TGTCTGAACCACTGGGTGGAAACTTTTGAAGAAATACACCCAGCTTGC 150793
117 ..... 117

```

150792 AGGTATACGGGAAATAGCGAGGAGTAGAATCTTTGTTGACAGTGT 150743
 117 117
 150742 CCTGGGTTGCCAGATATAGGGTAGAGCCTTGAGAAGTACGTGAAGAGT 150693
 118 ArgGlcysGlyLeuTyrPleuSerLeuGly 127
 150692 TGTCACTGTCCTTCCCCCAGGGGAATGTGACTGTGGCTGCTTGGGT 150643
 128 GlyPheHisGluArgGlyGlnAspTyrGluGlnThrGlnIleTyrAs 144
 150642 GGGTTCCATGAGGCGTGCCCAAGACTGGGAGCAGACTCGAATAATCTACAA 150593
 144 nCysHisValLeuLeuAsnSer..Lys..... 152
 150592 TTGTCACTGTCCTCTGACACAGCAAGGTGAGACTTATAACCTTTAGCC 150543
 152 152
 150542 TGGCTCTCCCATGCTCTTCTACCTAGATTCTCAGAAATTGTTCTCAAC 150493
 152 152
 150492 TCTTATTTCTTGACCCAGAGATTAGGGGTGCTCTACTTCACTTCTTA 150443
 152 152
 150442 GCCTATTAACATCTCTCTCTGGGAGAGTAGCAAGCTTACAGACA 150393
 153 GlyAlaValAlaIleThr 159
 150392 CCAGCACTGATATCTCTTCTTCTTCTACTGTAGGGGAGTAGTGCCACTT 150343
 159 yTrpArgTyrThrHisLeuCysAspValGluIleProGlyGlnGlyPromet 175
 150342 ACAGGAAGACACTCTGTGTGACGTAGAGATTCCAGGGGAGGCGCTATG 150293
 176 CysGluSerAsnSerThrMetProGlyProSerLeuGlnSerProValSe 192
 150292 TGTGAAGAAGCACTACACAGCTGGGCCCAAGCTTGAGTCACTGTGACG 150243
 192 rTrpProAlaGlyLysIleGly..Leu..... 200
 150242 CACACCAGAGAGCAAGTAGAGTGTGAAGAAGATGAGGAGGGAACAG 150193
 200 200
 150192 GAATCTTTGAAGTGCAGTAGAGATAGAAAGCCCTAAGAGAGGGGTA 150143
 200 200
 150142 ATGGAATATGACTAGATGCTGTGACAAACAGAGCAAGAACTACTAAG 150093
 201 AlaValCysTyrAspMetArg 207
 150092 TAGGCTGTTTTCATTCACAGATTGCTGTAGCTGTCTGTGATGACATGCGG 150043
 208 PheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
 150042 TTCCCTGAACTCTCTGTGCTGCTGCAAGCTGGAGCAGAGATACTTAC 149993
 224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTyrGluVal. 240
 149992 CATACCTTCAGCTTTTGGATCCATTACAGGCCACGCACTGGGAGGTAA 149943
 241 ..Leu..... 241
 149942 GATGATGCCCTTTTAAACATAAGGCTTTTCTAATCATCTTCCCC 149893
 241 241
 149892 CCTTGGCCCTACACGTTAAATTCCTTCCCTTTCACACTAATGGAAC 149843

241 241
 149842 TCATTCGCCAGATATTCTCTCATGATAGTAAATAGTAATCATTC 149793
 241 241
 149792 TAGCAATTACCAAAATAGTACAAATGGGTAGATTGGTCTTAATGCC 149743
 241 241
 149742 TCACCTGTCACTTCCCACTATTGCTACTGTACTTAACTGAACACA 149693
 242 LeuArgAlaArgAlaIleGluThrGlnCysTyr 252
 149692 TCTCATGCCAGGTGTGTGCGGGCCCGTATCGAATCCAGTGTCTA 149643
 252 rValValAlaAlaAlaGlnCysGlyArgHisGluLysArgAlaSer 269
 149642 TGTAGTGGCAGCAGACAGTGTGACGCCACCATGAGAAAGAGCAAGTT 149593
 269 yTrpGlnHisSerMetValAlaAspProTyrGlyThrValValAlaArgCys 285
 149592 ATGGCCACACGATGTGTGAGACCCCTGGGGAACAGTGTGGCCGCTGC 149543
 286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArg 302
 149542 TCTGAGGGGCGCAGCCCTCTGCTGGCCGAAATAGACTCAATATCTGCG 149493
 302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeu 319
 149492 ACAGTTGCGCGCCAGCACCTGCTGTTCAGCACCGCAGCGCTGACCTCT 149443
 319 yTrpGlnAsnLeuGlnHisProLeuSer 327
 149442 ATGGCAATCTGGGTCAACCACTGTCT 149417
 seq_name: gb_pr:AL591806
 seq_documentation_block: 200822 bp DNA linear PRI 30-JAN-2002
 LOCUS AL591806
 DEFINITION Human DNA sequence from clone Rpl1-544M22 on chromosome 1, complete
 sequence.
 ACCESSION AL591806
 VERSION AL591806.16 GI:18476709
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 200822)
 AUTHORS Harrison,E.
 TITLE Direct Submission
 JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humbry@esanger.ac.uk Clone requests: clonequest@sanger.ac.uk
 On Feb 1, 2002 this sequence version replaced gi:17902927.
 COMMENT
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M3 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> Rpl1-544M22 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pACE3.6

This sequence is the entire insert of clone Rpl1-544M22. The true left end of clone Rpl1-137A12 is at 156538 in this sequence. The true right end of clone Rpl1-381D2 is at 145015 in this sequence.

FEATURES

Location/Qualifiers

1..200822

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="Rpl1-544M22"

/clone_lib="RPCI-11.2"

10395..10397

/note="Sequence from overlapping clone Rpl1-381D2 (AL162592). Assembly confirmed by restriction digest."

44098..44107

/note="Sequence from overlapping clone Rpl1-381D2 (AL162592). Assembly confirmed by restriction digest."

59585..59604

/note="Sequence from overlapping clone Rpl1-381D2 (AL162592). Assembly confirmed by restriction digest."

90797..90883

/note="Single clone region. Reads generated from a transposon library derived from a single PUC clone. Restriction digest data confirm the assembly."

90861..90883

/note="Sequence from uni-directional dGTP big dye terminator reads only."

BASE COUNT

55215 a

49052 c

46183 g

50372 t

ORIGIN

alignment_scores:

Quality: 1468.50

Ratio: 4.505

Percent Similarity: 49.394

Length: 660

Gaps: 8

Percent Identity: 48.939

alignment_block:

US-09-357-675c-21 x AL591806

Align seg 1/1 to: AL591806 from: 1 to: 200822

2

LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysPr 18

172789 CTGGGCTTCATCACCAGGCTCTCTACAGATTCCTCTGTCCTC 172838

18

OGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32

172839 TGGAGTCCGGATACCTCACTCACTGACTTGTGTGTCACGCCAGGTAA 172888

32

ACGTTTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 32

172889 ACCTTTTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 172938

32

ATGATATGTGGAGACACAGAGTGTCACTCACTCACTCACTCACTCA 172988

32

GAGCCCTACTAGCCCTGGGTCAAGTGCCTGTGAAGCATGATCAAAAG 173038

32

GAGTCCAGCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 173088

173039

GAAGTCCAGCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 173088

173088

173088

173088

173088

173088

173088

173088

173088

173088

173088

173088

33Argpr 34

173089 CTTCCTGTGCTATGAAATCTGAGATCTGCTATGCTTCTACAGGCC 173138

34

oatgAlaMetAlaIleSerSerSerSerSerSerSerSerSerSerSer 51

173139 CAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 173188

51

aLcysGlnValThrSerThrProAspLysGlnGlnAsnAsnAsnAsn 67

173189 TGGGCCAGTAAATCATGACCCAGACCAAGACAGAACTTTAAACAT 173238

68

AlaGlnLeuValArgGlnAlaIleArgLeuGlnValAsnLeuAlaPhe 84

173239 GCTGAGCTGGTGGAGAGGCTGCGAGCTGGTGGCTGGCTGGCTGG 173288

84

uProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThrLeuHis 101

173289 GCCTGAGGCACTTGTACTTCTTGTGACGGGACCTGCGAGAACCTAC 173338

101

eUeSerGluProLeuGlnGlyLysLeuLeuGlnGlnGlnGlnGlnGln 117

173339 TGTCTGAACCACTGGGTGGGAACTTTGGAGAATACACCCAGCTTGC 173388

117

AGGTATCAGGAAATATGACGAGGAGAGGTAGAACTTTGTTGACAGTGT 173438

117

CCCTGGGTTGCCAGATATGAGGAGAGCCTTGAGAACTAGTGAAGAGT 173488

118

ATGTCAGTGTCCCTTCCCTCCAGGAAATGTGAGCTGTGCTGTGGGT 173538

128

GlyPheHisGlnArgGlnGlnAspTrpGlnGlnThrGlnIleTyrAs 144

173539 GGTTCATGAGCGGTGGCAAGACTGGAGGACGACCTGAGAAATCTCA 173588

144

nCysHisValLeuLeuAsnSer..Lys..... 152

173589 TTGTACAGTGTGCTGTGACAGCAAGGTGAGACTTTTAACCTTTAG 173638

152

TGCCTTTCCCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 173688

152

TCTTATTTCTTGAACCAAGGATTTAGGGGTGTCTCTCTCTCTCTCTCT 173738

152

GCTTATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 173788

153

CAAGCACTGATATTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 173838

159

YrArgLysThrHisLeuCysAspValIleProGlnGlnGlnGlnGln 175

173839 ACAAGGAAGACACATCTGTGACGTAGAGATTCAGAGGAGGAGGCT 173888

176

CysGlnSerAsnSerThrMetProGlnProSerLeuGlnSerProVal 192

173889 TGTGAAGCAACTCTACCATGCTCTGGGCGCAAGTCTTGAAGTACCT 173938

192

rThrProAlaGlyLysIleGly..Leu..... 200

173939 CACACCAAGCAGGAGGATAGAGTGTGAAAGATGAGGAGGAGGAGAG 173988

```

200 ..... 200
173989 GAATACCTTTGAACGAGCAGTAGAGATAGAAAGCCCTAAGAGAGGGGTA 174038
200 ..... 200
174039 ATGGAAATATGACTAGATGCTGTGACAAACAGACGAGAACTACTAAG 174088
201 ..... AlavalcystyrAspmetarg 207
174089 TAGGCGTTTTCATTCAGATTGCTGTAGCTGTCTGTATGATGACGG 174138
208 PheProGluLeuSerLeuAlaLeuAlaGluAlaGluLeuLeu 224
174139 TTCCCTGAACCTCTCTGCGCATTTGGCTCAAGCTGAGACAGATACTTAC 174188
224 TyrProSerAlaPheGlySerIlePheGlyProAlaHisTrpGluVal 240
174189 CTATCCTTCAAGCTTTTGGATTCATTACAGGCCCGCCACTGGAGGTAA 174238
241 Leu..... 241
174239 GATGATGCGCTTTTAAACATAAGGCGCTTTTCTTAACCTCATCTCC 174288
241 ..... 241
174289 CCTTGCCCTACACAGTTAAATCCCTTCCCTTCCACCTAATGGAAAC 174338
241 ..... 241
174339 TCATTCGCCAGATATTTCTCTCATGAATAGTAAATTAATCAATGCC 174388
241 ..... 241
174389 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGCC 174438
241 ..... 241
174439 TCACCTGTACCTCCCACTAATTTGCTACATGTAATTAAGTAACACACA 174488
242 ..... LeuArgAlaArgAlaIleGluThrGlnCys 252
174489 TCTCATGCCAGGTGTTGCTGGGGCGCGTGTATGAAACCCAGTGCTA 174538
252 rValValAlaAlaAlaGlnCysGlyArgHisGlnCysArgAlaSer 269
174539 TGTAGTGGACAGACAGCAGTGTGGACGCCACCATGAGAAAGACAAATT 174588
269 yTGlyHisSerMetValValAspProTrpGlyThrValAlaAlaArgCys 285
174589 ATGGCCACAGCATGTGTGTAGACCCCTGGGAAACAGTGGTGGCCCTGC 174638
286 SerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArg 302
174639 TGTAGAGGGCCAGGCGCTGTGCTGCCAATAGACTCAACTATGTGG 174688
302 GlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeu 319
174689 ACAGTTGGCCGACACACTGCTGTGTTCACAGCAGCCGAGCCGACTCT 174738
315 yTGlyAsnLeuGlyHisProLeuSer 327
174735 ATGGCAATCTGGGTCAACCACATGTCT 174764
seq_name: gb_htg:AC084821
seq_documentation_block: 211771 bp DNA linear HTG 29-JAN-2002
LOCUS AC084821 Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE,
DEFINITION 1 ordered pieces.
ACCESSION AC084821
VERSION AC084821.24 GI:18390260
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

```

```

SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
1 (bases 1 to 211771)
Deschamps, S., Gu, W. and Roe, B.A.
TITLE
Mus musculus BAC Clone rp23-395h6
JOURNAL
Unpublished
REFERENCE
AUTHORS
2 (bases 1 to 211771)
Deschamps, S., Gu, W. and Roe, B.A.
TITLE
Direct Submission
JOURNAL
Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Jan 29, 2002 this sequence version replaced gi:16076995.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKMOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 211771: contig of 211771 bp in length.
1 Location/Qualifiers
1. 211771
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-395h6"
/clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT 55622 a 48688 c 49207 g 58231 t 23 others
ORIGIN
alignment_scores:
Quality: 1105.50 Length: 676
Ratio: 3.893 Gaps: 8
Percent Similarity: 42.012 Percent Identity: 37.722
alignment_block:
US-09-357-675C-21 x AC084821 ..
Align seg 1/1 to: AC084821 from: 1 to: 211771
29 CysAlaGlnProArgProAlaMetAlaIleSerSerSerCysG1 45
|||||: ||||||:|||||: ||||||:|||||: ||
975 TGTGTATTCACAGGCGCCAGAACATGTC...TGTACTATTCCTGGGA 1021
45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnG 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1022 GCTGCCCGCTGGCGTGTGCTGCAGGTATACATCAACACCAACAAAGCA 1071
62 lAsnPhelysThrcysAlaGlnLeuValArgGluAlaAlaArgLeuGly 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1072 AGAACTTTAAACATGTGCTGAGTTGGTTCAGAGGCTCCAGACTGGGT 1121
79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspR 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1122 GCTTGCTGGCGCTTTCTGCTGAGGCAATTTGACTTTATTTGCCAGAAACC 1171
95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuLeuGluG 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1172 TCCCGAGACATTAATCCTGTCGGAACCACTGAATGGGATCTTTGGGCC 1221
112 lutyThrGlnLeuAla..... 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

[illegible]

seq_documentation_block: 4481 bp DNA linear ROD 23-JUL-1998

LOCUS AF069985

238 238

2172 TCCAGACACTTCAGTGATAGCCACAGGTTCMAAGCAGCATGGCTTGAC 2221

238 238

2222 TAAAAAGAAAAAATAAATAGAACTGGGCACTGGGGGTGCACACTTTAA 2271

238 238

2272 TCCACGACACTCAGGAGCAGAAAGCAGGTGATCTCTGAGTTTGAGGCCAG 2321

238 238

2322 CCTGGTTCACAGAGTGAGGTTCCAGGACAGCCAGGGCTACACAGAAACC 2371

238 238

2372 CTGCTCAAAAACAAAAAGCACCACACTTCAAAAGATGGGGCATGG 2421

238 238

2422 TGGTGATGCTCTTCATCTCAGCACTGTGAGGCAAAAGCAGGAGATCTC 2471

238 238

2472 TGTGGGCTTAAGCCAGCCAGGGTTACACAGGAGACCTTATCTCAAAAGT 2521

238 238

2522 ACAAACACAAACACTCCTCCTTCTCTACTTATTTCTCTCTTTCATCT 2571

238 238

2572 GGTGAGGAATTCATTCTTAGCTAGTGTGATCATGTTAAAGATTAG 2621

238 238

2622 TTATTAACTAGTTGTTAGTAGAAGATTGGCCACAGTCCCGCTT 2671

238 238

2672 CTGTGACCTGCGTTTAAAGCGCTGATGTTCTGGAGTCAACGCACTCT 2721

239 Glu.ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVa 253

2722 CATCTCAGAGTCTGCTGCGGGCCGCGCCACTGTAATCTCAGTCTATGT 2771

253 ValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgLysLeuArg 270

I:::||||| 2772 AATGACGACGCGCAGTGTGAGCCGACCATGAACAACAGACAAAGTATG 2821

270 IyHisSerMetValAlaAspProTrrGlyThrValValAlaArgCysSer 286

2822 GCCATAGCATGTGTTGATGCCGTGGGACACAGTGTGGCCGCTGCTTC 2871

287 GlnGlyProGlyLeuCysLeuAlaArgLleAspLeuAsnTyrLeuArgG 303

2872 GAGGACCGAGGCTCTGGCTTGGTTCGAATTGATCTCCACTTTCACACA 2921

303 IleuArgArgHisLeuProValPheGlnHisArgArgProAspLeuArg 320

2922 GATGGCCACACACTGCTGTGTTCACACACCGCAGACTGACTTATG 2971

320 IyAsnLeuGlyHisProLeuSer 327

2972 GCAGTCTGGGTATCCACTCTCT 2994

seq_name: gb_ro:AF069985

DEFINITION	Mus musculus nitrilase homolog 1 (Nlt1) gene, alternatively spliced product, complete cds.
ACCESSION	AF069985
VERSION	AF069985.1
KEYWORDS	GI:3242979
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 4481) Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tilib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
TITLE	2 (bases 1 to 4481) Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tilib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
JOURNAL	Direct Submission
FEATURES	Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. 38th St., Philadelphia, PA 19107, USA
SOURCE	Location/Qualifiers 1..4481 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="1" /map="1q21-q23" join(541..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3944) /gene="Nlt1" /note="alternatively spliced" /product="nitrilase homolog 1" 541..3944 /gene="Nlt1" 541..606 /gene="Nlt1" /note="alternatively spliced" /number=1 join(605..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3658) /gene="Nlt1" /note="alternatively spliced" /codon_start=1 /product="nitrilase homolog 1" /protein_id="AAC40184.1" /db_xref="GI:3242980" /translation="MIGFTTRPHDLCGYRLRIPVLCTOPRPTSMSSSTSWELPLVAVCOVSTPNKQENFKTCALVQEARAGACLAFLPEAFDEFIARNPASTLISEPLNGLDGYYSOLAREGIMVLSGFHEHGOENOKLYNCHVILNSGVAVSRKTHLGDVEIPGGPMRESNYTKRGTLPKPPVKKRGVGLATCYDMRPPLSLKLAQCAEILTPYRSFSSVTGPAHWEVLLRRAIESOCYVLAACGGRHETRASTGSMVYDPWGTVVARSSEPGCLARIDHFLQOMRONHLPVQHRPDLXSLGHPLS" 863..986 /gene="Nlt1" /note="alternatively spliced" /number=2 1177..1263 /gene="Nlt1" /number=3 1638..1889 /gene="Nlt1" /number=4 2015..2118 /gene="Nlt1" /number=5 2362..2495 /gene="Nlt1" /number=6 2626..2751 /gene="Nlt1"
gene	
exon	
CDS	
exon	
exon	
exon	
exon	
exon	

exon		/number=7 3392..3944 /gene="Ntli"			
BASE COUNT	1101 a	1132 c	1133 g	1115 t	
ORIGIN					
alignment_scores:					
	Quality: 1102.50		Length: 680		
	Ratio: 3.668		Gaps: 8		
Percent Similarity:	41.912		Percent Identity:	37.353	
alignment_block:					
US-09-357-675C-21 x AF069985 ..					
Align seg 1/1 to: AF069985 from: 1 to: 4481					
29	CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysG1	45			
1624	TGCGTATTTCACAGGGCCCAACCATCTCC...TCATCAACTCTCTGGGA	1670			
45	uLeuProLeuValAlaValAlaCysGlnValIThrSerThrProAspLysG1	62			
1671	GCTGCCCCCTGGTGGCTGTGTGCCACGGAATCATCAACCAACACAGCAG	1720			
62	hAsnPhelYThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly	78			
1721	AGAACTTTAAACATGATGCTGAGTGGTCAAGAGCGCTGCACACTGGGT	1770			
79	AlaCysLeuAlaPhelLeuProGluAlaPhasPheIleAlaArgAspR	95			
1771	GCTGGCTGGCCCTTCTGTCCGTGAGGCACTTTCATTTGACAGAAACC	1820			
95	oAlaGluThrIleuHisLeuSerGluProLeuGlyLysLeuLeuG1	112			
1821	TGCCGACACTTACTCTGTGCCACACACTGAATGGGATCTTTGGGCC	1870			
112	LuIuYrThrgInLeuAlaArg..Glu.....	119			
1871	AATATGCCACGCTTGGCAGGTACAAAGGTAAGGTTGGGGGGGGGGGT	1920			
119	119			
1921	TGTGTGTGAGAACTCCCTGTGTGAGCAGTGTCTGAATTGCCACAGAGG	1970			
120	120			
1971	GGGTAGAGTTGT	2020			
120	seGLeuTrpLeuSerLeuGlyLysPheHisGluArgGlyGlnAspTrpG	137			
2021	TGGAACTCTGGCTCTCTTGTGGCGGTTCCAGAGCGTGGCCAAACATGGG	2070			
137	LuGlnThrgInLysIleYrAsnCysHisValLeuAsnSerLys..	152			
2071	AGCAGATCAGAAATCTACAATTGCATGTGCTTTGAACACAGAAAGGT	2120			
152	152			
2121	GAGACGTCAATAGCCCTTACTGTGCTCTTCCAGGCTTCTTACCTTTC	2170			
152	152			
2171	CTCAACTCAGATTTCAGCATTCTTCTCAACTGCTTGGCCTTGGGGTT	2220			
152	152			
2221	TTATGCTTTGTTCTCAGCCTACAGATTACTCTTGGGAGAGAGTAAG	2270			
152	152			
2271	CTGAGCTGTAGAGTCTGTACTGTGATATTTCTTCTGTCTCTCTCTC	2320			

```

238 ..... 238
3221 TCCCTTCATCTGGTGAGGAAATTCATTCTTAGCTAGGTGCATCATG 3270
238 ..... 238
3271 TTAAAGAGTTAGTTATTACTCTAGTGGTTAGTGGAAGATTGGCCCA 3320
238 ..... 238
3321 CAGTGGCCGCTTCTGTCTACTTGCCATTAAAGCTGTATGTTCTGAGT 3370
239 ..... Glu.ValIeuLeuAryAlaIleGlu 249
3371 CAACGCACGCTCATCTCCTCAGGTGCTGCGGGCCCGGCCCATTCGAAC 3420
249 rGInCysTyrValAlaIAlaIAGlnCysGlyArgHisIleSgLUysA 266
3421 TCAGTGTCTATGTTAATAGCACACCGCAGGTGGACCCGCCATGAACAA 3470
266 rGAlaSerTyrGlyHisSerMetValAlaAspProTrrpLysThrVal 282
3471 GAGCAAGTATGAGCCATAGCATGTGGTGGACCCGGGGCAGACAGTGG 3520
283 AlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 299
3521 GCCCGCTGCTCCGAGGACAGGCCCTCTCCCTGCTGCAATTCATCTCCA 3570
299 nTyrLeuArgGlnIleuArgHisIleuProValPheGlnHisArgArg 316
3571 CTTTCTACACAGATGCGCAACCTGCTGCTGTTCCAGCACCGCAGAC 3620
3621 CTGACCTCATGCGACACTGCGGTCACTCACTCTCT 3655
seq_name: gb_htg:AC087229
seq_documentation_block:
LOCUS AC087229 213203 bp DNA linear HTG-05-FEB-2002
DEFINITION Mus musculus chromosome 1 clone rp23-191a19, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
ACCESSION AC087229
VERSION AC087229.15 GI:18497119
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Jiang,X., Song,L., Gu,W. and Roe,B.A.
TITLE Mus musculus Chromosome 1 BAC Clone rp23-191a19
JOURNAL Unpublished
2 (bases 1 to 213203)
AUTHORS Jiang,X., Song,L., Gu,W. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```
* * 1 11373: contig of 11373 bp in length
* 11374 11473: gap of unknown length
* 11474 43477: contig of 32004 bp in length
* 43478 43577: gap of unknown length
* 43578 126526: contig of 82949 bp in length
* 126527 126626: gap of unknown length
* 126627 213203: contig of 86577 bp in length.
Location/Qualifiers
1. 213203
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-191a19"
BASE COUNT 57858 a 50465 c 49150 g 55425 t 305 others
ORIGIN

alignment_scores:
Quality: 1083.00 Length: 676
Ratio: 3.813 Gaps: 9
Percent Similarity: 42.012 Percent Identity: 37.574

alignment_block:
US-09-357-675c-21 x AC087229 ..

Align seg 1/1 to: AC087229 from: 1 to: 213203

29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysG1 45
|||||:|||||:|||||:|||||:|||||:|||||:
48642 TGTGATTCACAGGCCACAGACCATGCTCC...TCATCAACTTCCTGGGA 48688
|||||:|||||:|||||:|||||:|||||:|||||:
45 uleuProLeuValAlaValCysGlnValThrSerThrProAspLysGln 62
|||||:|||||:|||||:|||||:|||||:|||||:
48689 GCGGCCCTGGTGGCTGTGTGCCAGGTAACATCAACACCAACAAACAGCAG 48738
|||||:|||||:|||||:|||||:|||||:|||||:
62 lnaAsnPhelysthrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 78
|||||:|||||:|||||:|||||:|||||:|||||:
48739 AGCACTTTAAACATGCTGTGAGTTGCTTCAGAGGCTGCCACACATGGGT 48788
|||||:|||||:|||||:|||||:|||||:|||||:
79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 95
|||||:|||||:|||||:|||||:|||||:|||||:
48789 GCTTGCCTGGCCCTTCTGCTGAGGCAATTTGACTTATTGCACGAAACC 48838
|||||:|||||:|||||:|||||:|||||:|||||:
95 oAlaGluThrLeuHisLeuSerGluProLeuGlyLysLeuLeuGluG 112
|||||:|||||:|||||:|||||:|||||:|||||:
48839 TGGCGAGACATTATCTCTGTCCGAAACACATGAGATGGGATCTTTGGGCC 48888
|||||:|||||:|||||:|||||:|||||:|||||:
112 lntYrThrGlnLeuAlaArg.Glu..... 119
|||||:|||||:|||||:|||||:|||||:
48889 AATATAGCCAGCTTGGCAGGTACAAAGGTAGGGTTGGGGGGGGGGTT 48938
|||||:|||||:|||||:|||||:|||||:|||||:
119 ..... 119
48939 GTGTGTGTAGAACTCCCTGTGTAGCAGTGTCTCTGATTGCCACAGAGG 48988
|||||:|||||:|||||:|||||:|||||:|||||:
120 .....CysGlyLe 122
|||||:|||||:|||||:|||||:|||||:|||||:
48989 GGTAGAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 49038
|||||:|||||:|||||:|||||:|||||:|||||:
122 uTrpLeuSerLeuGlyGlyPheHisGluArgGlyLnaAspTrpGluGln 139
|||||:|||||:|||||:|||||:|||||:|||||:
49039 CTGGCTGTCTCTGGGGGTTTCCACGAGCGGTGGCCAAAGACTGGGACACGA 49088
|||||:|||||:|||||:|||||:|||||:|||||:
139 hrcGlnIstIleTyrAsnGlyHisValLeuLeuAsnSerLysGlyAla.V 155
|||||:|||||:|||||:|||||:|||||:|||||:
49089 ATCAGAAATATCAATGTGCTTTTGAACAGACAGAGGTTGAGACG 49138
|||||:|||||:|||||:|||||:|||||:|||||:
155 a1..... 155
|||||:|||||:|||||:|||||:|||||:|||||:
49139 TCATAGCCCTTACTTGTCTTCCACAGGCTTCTTACCTTTCTCAACT 49188
|||||:|||||:|||||:|||||:|||||:|||||:
155 ..... 155
```

```
49189 CAGATCTTCAGCATTTCTTCTTCACACTGTGGCTTGGGGTTTATGCT 49238
155 ..... 155
49239 TCATGTTCTACGCTACAGATTACTTTGGGAGAAAGCTAAGCTGGACT 49288
155 ..... 155
49289 TGTAGAGTGTCTACTGAGATTTTCTCTCTCTCTCTCTCTCTCTCT 49338
156 .....ValAlaThrTyrAr 160
49339 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 49388
160 gLysThrHisLeuCysAspValGluIleProGlyGlnGlyPrometCysG 177
|||||:|||||:|||||:|||||:|||||:|||||:
49389 GAAGACACATCTGTGGATGTAGAGATCCAGGTCCAGGGGCCGATGAGAG 49438
177 luserAsnSerThrMetProGlyProSerLeuGluSerProValSerThr 193
|||||:|||||:|||||:|||||:|||||:|||||:
49439 AAGCAACTATATACCAAGCTGAGAGCAGCTCTTGAGCCACTGTCAAGACA 49488
194 ProAlaGly.Lys..... 197
|||||:|||||:|||||:|||||:|||||:|||||:
49489 CCGCTGGCAAGGTGGAGTTTGAAAGAAATGGAGACTATGTGAAAG 49538
197 ..... 197
49539 TGGTAAACCCCCCATTTGGTGCAGGCTAATGAAGCAGTCTCTCTCT 49588
198 .....IleGlyLe 200
49589 TGAATGGAGCAGGAGGAGATGATGAGGTGTGTCTCATTTTCAGTTGGTCT 49638
200 uAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
|||||:|||||:|||||:|||||:|||||:|||||:
49639 AGCAACTGTGTAGACATGGGTTCCCTGAACCTTTTGAATTTGGTCC 49688
217 lnaAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233
|||||:|||||:|||||:|||||:|||||:|||||:
49689 AAGCTGGGGCAGAAATATCTTACTTATCTTCACGCTTTGGATGTGTACA 49738
234 GlyProAlaHisTrp..... 238
|||||:|||||:|||||:|||||:|||||:|||||:
49739 GGTCCAGCCCACTGGGAGGTAGAGAAATGCTCTCATATATAGGCC 49788
238 ..... 238
49789 TCCTTTAACCCTTACTTTTATGTAAGTGGAGCAGAGTGTACAAACCTTAAT 49838
238 ..... 238
49839 CCAAGCACTTCAGTGAATAGCCACAGTTCAAAGCCAGCATGCTTGACT 49888
238 ..... 238
49889 AAAAAAAGAAATAGAAAGCTGGGAGTGGGGGTGCACACCTTTAAT 49938
238 ..... 238
49939 CCGAGCACTCAGAGAGCAGAAAGAGTGTGATCTGTAGTTTGAGGCCAGC 49988
238 ..... 238
49989 CTGGTTACAGAGTGGGTTCCAGAGACAGGAGGCTACACAGAAACC 50038
238 ..... 238
50039 TGTCTCAAAAAAAGCAAAAGCAACCCACACTTCAAGAGTGGGGCATGGT 50088
238 ..... 238
```

```

50089 GGTGCATGCTCTTCATCTCAGCACTGTGAGCAAGGACGAGATCTCT 50138
238 ..... 238
50139 GTGGGCTTAAGCAGCCAGGGTTACACAGGAGACCTTATCTCAAAAGTA 50188
238 ..... 238
50189 CAAACACAAACACTCTCTCTTCTACTTGTATTTCTTCTTTCATCTG 50238
238 ..... 238
50239 GTGAGGAATTCATTCTCTAGTAGTGTAGTCATGTAAAGAGTTAGT 50288
238 ..... 238
50289 TATTACTAGTTGGTTTACTGAGAGATTGGCCACAGTCCCGCTTC 50338
238 ..... 238
50339 TTGTCACTTGCCTATTAAACGCTGTATGTTCTGGAGTCAACGACGCTCTC 50388
239 ..... 50438
50389 ATCCTGAGTGTCTGTGCGGGCCGCCATTGATCTCAGTGTATGTA 50438
254 VALAIAAIAAGINCYSGLYARGHSHISGLULYSARGALASERTYRG1 270
50439 ATAGCAGCAGCAGCGAGTGTGAGCCACCAAGCAAGCAAGTATTATG 50488
270 YHLSerMeValValAspProTrpGlyThrValValAlaArgCysSerG 287
50489 CCATAGCATGTGTGTGACCCGTTGGGACACAGTGTGCGCCGCTGCTCCG 50538
287 InGlyProGlyLeuGlyLeuAlaArgIleAspLeuAsnTYrLeuArgIn 303
50539 AGGACACAGCGCTCTGCTGCTCGATGATCTCCACTTCTTCAACAG 50588
304 LeuArgArgHisLeuProValAlaPheGlnHisArgArgProAspLeuTYrG1 320
50589 ATCGCGCAACACCTGCTGTGTTCCAGACCGCAGACCTGACTTATG 50638
50639 CAGTCTGGTCACTCCTCTCT 50660
seq_name: gp_htg:AC105589
seq_documentation_block:
LOCUS AC105589 178418 bp DNA linear HTG 09-JAN-2002
DEFINITION Rattus norvegicus clone CH230-242B2, *** SEQUENCING IN PROGRESS
ACCESSION AC105589
VERSION AC105589.1 GI:18092812
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 178418)
Muzny,D.M., Adams,C., Adio-Oduola,B., Al-Osman,F.R., Allen,C.,
Alstrooms,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbala,J.,
Benton,J., Blincoe,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

```

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

COMMENT

```

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,N., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivel,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korhavi,S.,
Kovar,C., Kretovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W.,
Louiseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S.,
Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshari,N.,
Stinson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 178418)
Morley,K.C.
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNGI
Center clone name: CH230-242B2
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findhaplolist
Consensus quality: 132588 bases at least Q40
Consensus quality: 142233 bases at least Q30
Consensus quality: 152072 bases at least Q20
Estimated insert size: 141845; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
8303: contig of 8302 bp in length
8403: gap of unknown length
8403: contig of 10070 bp in length
18473: gap of unknown length
18573: gap of unknown length
18573: contig of 6350 bp in length
24922: gap of unknown length
24922: contig of 5877 bp in length
30899: gap of unknown length
30900: contig of 7321 bp in length
31000: gap of unknown length
38321: gap of unknown length
38420: gap of unknown length

```

```
* 38421 43244: contig of 4824 bp in length
* 43245 43344: gap of unknown length
* 43345 48228: contig of 4884 bp in length
* 48229 48328: gap of unknown length
* 48329 54595: contig of 6267 bp in length
* 54596 54695: gap of unknown length
* 54696 58551: contig of 3856 bp in length
* 58552 58651: gap of unknown length
* 58652 62204: contig of 3553 bp in length
* 62205 62304: gap of unknown length
* 62305 66587: contig of 4283 bp in length
* 66588 66687: gap of unknown length
* 66688 70928: contig of 4241 bp in length
* 70929 71028: gap of unknown length
* 71029 75627: contig of 4599 bp in length
* 75628 75727: gap of unknown length
* 75728 78272: contig of 2545 bp in length
* 78273 78373: gap of unknown length
* 78373 80301: contig of 1929 bp in length
* 80302 80401: gap of unknown length
* 80402 83675: contig of 3274 bp in length
* 83676 83775: gap of unknown length
* 83776 87173: contig of 3398 bp in length
* 87174 87273: gap of unknown length
* 87274 90196: contig of 2923 bp in length
* 90197 90296: gap of unknown length
* 90297 92165: contig of 1769 bp in length
* 92066 94612: gap of unknown length
* 92166 94712: contig of 2447 bp in length
* 94613 94713: gap of unknown length
* 94713 98755: contig of 4043 bp in length
* 98756 98855: gap of unknown length
* 98856 102543: contig of 3688 bp in length
* 102544 102643: gap of unknown length
* 102644 105833: contig of 3190 bp in length
* 105834 105933: gap of unknown length
* 105934 108499: contig of 2366 bp in length
* 108500 108599: gap of unknown length
* 108600 111672: contig of 3073 bp in length
* 111673 111772: gap of unknown length
* 111773 114662: contig of 2890 bp in length
* 114663 114762: gap of unknown length
* 114763 116145: contig of 1383 bp in length
* 116146 116245: gap of unknown length
* 116246 118147: contig of 1902 bp in length
* 118148 118247: gap of unknown length
* 118248 120579: contig of 2332 bp in length
* 120580 120679: gap of unknown length
* 120680 122618: contig of 1939 bp in length
* 122619 122718: gap of unknown length
* 122719 124539: contig of 1821 bp in length
* 124540 124639: gap of unknown length
* 124640 126651: contig of 2012 bp in length
* 126652 126751: gap of unknown length
* 126752 128460: contig of 1709 bp in length
* 128461 128560: gap of unknown length
* 128561 130592: contig of 2032 bp in length
* 130593 130692: gap of unknown length
* 130693 132864: contig of 2172 bp in length
* 132865 132965: gap of unknown length
* 132966 134881: contig of 1917 bp in length
* 134882 134981: gap of unknown length
* 134982 136682: contig of 1701 bp in length
* 136683 136782: gap of unknown length
* 136783 139138: contig of 2356 bp in length
* 139139 139238: gap of unknown length
* 139239 140757: contig of 1519 bp in length
* 140758 140857: gap of unknown length
* 140858 142558: contig of 1701 bp in length
* 142559 142658: gap of unknown length
* 142659 144303: contig of 1645 bp in length
* 144304 144403: gap of unknown length
* 144404 145490: contig of 1087 bp in length
```

```
* 145491 145590: gap of unknown length
* 145591 147171: contig of 1581 bp in length
* 147172 147271: gap of unknown length
* 147272 148639: contig of 1368 bp in length
* 148640 148740: gap of unknown length
* 148740 150295: contig of 1556 bp in length
* 150296 150395: gap of unknown length
* 150396 151505: contig of 1110 bp in length
* 151506 151605: gap of unknown length
* 151606 152813: contig of 1208 bp in length
* 152814 153423: gap of unknown length
* 153424 155523: contig of 2310 bp in length
* 155524 155524: gap of unknown length
* 155525 156880: contig of 1357 bp in length
* 156881 156980: gap of unknown length
* 156981 158045: contig of 1065 bp in length
* 158046 158145: gap of unknown length
* 158146 159255: contig of 1110 bp in length
* 159256 159355: gap of unknown length
* 159356 160417: contig of 1062 bp in length
* 160418 160517: gap of unknown length
* 160518 161644: contig of 1127 bp in length
* 161645 161744: gap of unknown length
* 161745 163180: contig of 1436 bp in length
* 163181 163280: gap of unknown length
* 163281 164483: contig of 1203 bp in length
* 164484 164583: gap of unknown length
* 164584 166039: contig of 1456 bp in length
* 166040 166139: gap of unknown length
* 166140 167290: contig of 1151 bp in length
```

```
alignment_scores:
  Quality: 1059.00      Length: 835
  Ratio: 3.507          Gaps: 11
  Percent Similarity: 36.168  Percent Identity: 31.976
```

alignment_block:

US-09-357-675c-21 x AC105589 ..

Align seg 1/1 to: AC105589 from: 1 to: 178418

```
2 LengthPhenilethylarginineProphorohisArgphenylserLeucylsPr 18
|||||
39095 CTGGGCTTCATCCACGAGCCTCCATCACTCCCTCCCTGGTGGTAC 39144
18 oglyLeuArgIleProGlnIleSerValLeucylsAlaGlnPro..... 32
||| |||:|||||:|||||:|||||:|||||:|||||
39145 CGGATACCGATTACCTCAATCTCAGTACTTTGTACTCAGCCAGGTAC 39194
32 ..... 32
39195 ATATTTATGTTCTTGTGTACGACGTAAGTTCATGTCATGTTATTA 39244
32 ..... 32
39245 CCCGAATAGTAGCCTGCATCCACCAACATTAAATGCGTACTCCAG 39294
32 ..... 32
39295 TACCTTAGGTTATTAAATGCTCTGAGAGATGAAAAACAAGTCGTAGG 39344
32 ..... 32
39345 TTGGTGGACACACCTCTATCCACGACACTCGGGAACAGCAGAGAGAT 39394
32 ..... 32
39395 CAACTAGAGATTAGATCATCCCGCTACATAGAAAGTCTGCTGGGC 39444
32 ..... 32
39445 TACAAAAAGACCCGGTCTCAAGGGAGGTAATAATCAGAAAAAAG 39494
```



```
32 ..... 32
39495 TAGTTCCGGCTCCACCCCACTTGAGCCTCCCAAGTTCTCCGGGCA 39544
33 ..... ArgProArgAl 36
39545 CCGGAGTAATCTCGAANAACCTGCTGTGTCATTTCACAGGCCAGAGC 39594
36 AneIaIaIeSerSerSer...SerCysGluLeuProLeuValAlaValC 52
39595 CATGGCAGACTCCTCATCGACTTCCTGGAGCTGCCCTGGTGGCTGTGT 39644
52 yslInaValThrSerThrProAspLysGlnGlnAsnPhelysthrcysAla 68
39645 GCCAGGTAACTATACACCAACAACAGCAGAGACTTTAAACGTGTGCT 39694
69 GluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhelLeuPr 85
39695 GACCTGGTTCAGAGGCTACACAGACTGGGGCTTGCCTGGCTTCTGCC 39744
85 ocGluAlaPhelAspPheIleAlaArgAspProAlaGluThrLeuHisLeuS 102
39745 TGGAGCAATTTGACTTTATTCACGAAATCCTGCCGAGACATTTACTCCTGT 39794
102 erGluProLeuGlyGlyLysLeuLeuGluGluThrThrGlnLeuAla... 117
39795 CCGAACCACTGGATGGGACCTTTTGGGACAAATATAGCCAGCTTGGCCAG 39844
117 ..... 117
39845 TACAAGGATAGGGTGGGGTGGGGGAGGGGTANAATCCCTGTTGGGAG 39894
117 ..... 117
39895 TGTTCCTGATTCGCCAATAGAGGGGTAGAGTTGCATGTCCTCCCTCCC 39944
118 ..Arg..GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgI 133
39945 CTCAGGGAAATGGTGAATCTGGCTGTCTCTTGGTGGTTCACCAACCTGG 39994
133 yglAspTrpGluGlnThrGlnLysIleTyraAsnCysHisValLeuLeuAl 150
39995 CCCAGATGGGACAGAAATCCGAAATCTACATTTGCTCGGGCTTCTGA 40044
150 sn..SerLys..... 152
40045 ACAACCAAGGTGAGACATTCCTCAACTTGCTCTCTTAAGTGC 40094
152 ..... 152
40095 ATCTAACTTCCTCTCACTTAGAACTTAACTTTCTCTCACTGCTGGCT 40144
152 ..... 152
40145 TTAGGGGTTTTTTTGTGTTTGTGAATGCTCTGTTCTTAGCCTAAAA 40194
152 ..... 152
40195 TTACTCTTGGGAGAGAGCTAAGCTAGCTCCAGTGTCTGTAGTGCAGC 40244
153 ..... GylAlaValAlaIleThrTyraArgLysThrHisL 164
40245 ATTTTCTCTTACTTTAGGATCAGTAGTGGCAGTTACAGGAAAGCACATC 40294
164 euCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
40295 TATGTGATGTAGAGATCCAGGTCCAGGGGCTATGAGAGAAAGAACTAT 40344
181 ThMeTrProGlyProSerLeuGluSerProValSerThrProAlaGlyLy 197
40345 ACCATGCTGTGATACGCTTTTGAACCACTGTCAAGACACAGCAGAGCAA 40394
197 sileGly..Leu..... 200
40395 GGTAGAGTTTGAAGAAAGAAATAGGAATATGTTGAAGAAAGTTGAACATC 40444
200 ..... 200
40445 TCCCTTGTGTAGATGATGATGAGAGTAGATATTCANGTTCAGTTGACAG 40494
201 ..... AlaVal 202
40495 GACAGGAGATGATGAGTGGTGTGTTTCTCGTTTCAGGGTGGTCTAGCAATC 40544
203 CysTyraSphMeIaArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaG 219
40545 TGTATGACATGGGGTTCCTCGAATTTCTCTGAATGGGCTCAAGCTGG 40594
219 yAlaGluIleLeuThrTyraProSerAlaPheGlySerIleThrGlyPro 236
40595 GCCAGAAATACTTACTATATCTTACGCTTTGGATGTGTTACGGGCCAG 40644
236 IaHisTrpGluVal..... 240
40645 CCCACTGGGAGTAGAGGAAATGCCCTTCTTAACAATATAGCCCTCTTGA 40694
240 ..... 240
40695 ATCTTATTTTATGATGAGGGGACAGAGTAGACAAACCTAACTCTAAGC 40744
240 ..... 240
40745 ACTTCAGGGGATAGCCACAGGTTCAAGCCAGCATAGTATGCTCGAGT 40794
240 ..... 240
40795 AAAACAGAAACAAATAGAACTGGGCACTGGGGGTGCACACCTTTAAT 40844
240 ..... 240
40845 CCCAGCCTCAGAGAGGAGCAAGAGCAAGTGTGAGTTCAAGACAGCT 40894
240 ..... 240
40895 GGTCTACAGATAGTGTCCAGAGACGCGGCTACACAGAGAAACCGTC 40944
240 ..... 240
40945 TCAAAACAAACAAACAAACAAACAAAGTAACCCACACCAACCCAAA 40994
240 ..... 240
40995 ATAGTGTGATGTGTGTGTGTCTTTTATCTCCGCACTGTAAAGCAGA 41044
240 ..... 240
41045 GCGCGGAGATCTCTGTGGGTCCAGGCCAACAAAGTTATACAAGAGAGAC 41094
240 ..... 240
41095 CTTATATAAAAAATACAAAAACAAACACACACCCCTCCCTCTTACTGA 41144
240 ..... 240
41145 AGGATTTTCTCCCTCCGATCTGGTGTGAGAAATTCGTTCTTAGATAGG 41194
240 ..... 240
41195 TGTAGTCATGTATAGTTAGTTACTAATCTAGTGTATCATCAATATGCAAT 41244
240 ..... 240
41245 GAGTAGATGGGCCACAGTGTCCGCTTGTTCACACTGGCCTAATTAATG 41294
241 ..... Leu..LeuArgAla 244
41294 ..... 244
```

	41295	CNCATGTCCTCGATACAGCAGCTGCATCCATCCAGTGGTGCGTGGCGGCC	41344
	245	ARGAIAIIIEGLIUTHrGlncystYrValValAlaAlaAlaAlaGlncysGIyAr	261
	: : : : :	
	41345	CGGGCATTTGAGCTCAGTGTACTGATGGAAGCAGCAGCAAGTGTGGAGC	41394
	261	GHTSHSGILtYsArgAlaSerTyGIyHISserMatValAlaAspProT	278
	: : : : :	
	41395	CCACCATTGAGACAGAAGCAAGTTACGGCCACAGCAATGGTGGACCCCT	41444
	278	rPGLyThrValValAlaAlaArgcYsserGIUGlyLPProGIyleucysLeuaLa	294
	: : : : :	
	41445	GGGGCACAGTGGTGGCGACGCTTTCGAGGAGAACAGGCTTGGCTTGCT	41494
	295	ArGIleAspLeuAsnTYrLeuArgGlnleuArgARHISleuProValrh	311
	: : : : :	
	41495	CGAATGATCTCCACCTTTCTGCACACAGATGCCCAACACTTGGCTGTGT	41544
	311	eGlnHISArgArGProAspLeuTYrIGlyAsnleuGlnHISProLeuSer	327
	: : : : :	
	41545	CCAGCATCCACAGACCTGACCTTTATGGCAGTGC...CTGCCACCTGCC	41590
seq_name:	gb_ov:AF284575		
seq_documentation_block:			
LOCUS	AF284575	1214 bp	mRNA linear VRT 22-JUL-2000
DEFINITION	Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.		
ACCESSION	AF284575		
VERSION	AF284575.1 GI:9367117		
KEYWORDS	African clawed frog.		
SOURCE	Xenopus laevis		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;		
REFERENCE	Xenopodinae; Xenopus. 1 (bases 1 to 1214) Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J., Bieganski,P., Pekarsky,T., Croce,C.M. and Brenner,C. Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers Curr. Biol. 10 (15), 907-917 (2000)		
AUTHORS	2 (bases 1 to 1214) Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J., Bieganski,P., Pekarsky,T., Croce,C.M. and Brenner,C. Direct Submission Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA		
JOURNAL	200414396		
MEDLINE			
FEATURES	location/qualifiers 1..1214 /organism="Xenopus laevis" /db_xref="taxon:8355"		
SOURCE	1..1214 /gene="Nit1" 180..1046 /gene="Nit1" /note="Nit1, similar to the Nit domains of Caenorhabditis elegans NitFhit and Drosophila melanogaster NitFhit" /codon_start=1 /product="Nit protein 1" /protein_id="AAFG7104.1" /db_xref="GI:9367118" translation="MAGAHRIPLAVCOMSTSDKEKFAFCSRLIRAAARRACWFL PEADYIGGSTLEETLAESEHGDTIDRYQLAREGLMLSLGFEHKGNWDPOIT SNCHVVDNFTGHVSVRKAHLDPDVALQNVLSRESSSTLPAGALIRIPSPKGTL GVSGDLRPFESIALAQGAELLTYPASFLLTGLAEVLRLRALETQCYYVAAG TDRIENERTSYGMAMVDDPMGLIVGQQEGSTGICYAIDIPIEMRVBRDMPVMHRRT DIKVICSFNPMD"		
gene			
CDS			
BASE COUNT	281 a	305 c	345 g 282 t 1 others
ORIGIN			

	Quality:	968.50	Length:	307
	Ratio:	3.905	Gaps:	1
	Percent Similarity:	80.782	Percent Identity:	57.980
<p>alignment_block:</p> <p>US-09-357-675C-21 x AF284575 ..</p>				
<p>Align seg 1/1 to: AF284575 from: 1 to: 1214</p>				
16	LeucySPProGlyLeuAlaIleProGlnLeuSerValLeuCysAlaGlnPr	32		
144	CTGTCTCCGGATATATGGGGGTATGAGAACTTGAGTGCATG.....	182		
32	okgrProAlaMetAlaIleSerSerSerSerCysGluProLeuVal	49		
183GCTGGTGCCCAACAGCCCTGA	204		
49	AlaIleValCysGlnValThrSerThrProAspLysGlnGlnAsnPhel	65		
205	TTGGCGGTGGCCAGATGACTTCAACCTCTGTAAAGAAATTTGCC	254		
66	ThrCysAlaGluLeuValArgGlnAlaAlaArgLeuValLeuVal	82		
255	ACGTGTTCGGAGCTGATCCGGGAGGCTCCGGGCGTGCCTTCATGGT	304		
82	ArheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThr	99		
305	CTTTTGCCGGAGGCTTTGACTATATCGGGGCGAGCATTTGAGAGACGC	354		
99	eunHisSerLysGlnProLeuGlyLysLeuLeuGlnGluTrpThrGln	115		
355	TCAGTCTGAGTGAAGTCTCTACATGCGGACACCATTTACGCTTACACCAA	404		
116	LeuAlaIleArgLysGlyLeuThrPheLeuSerLeuGlyLysPheHisGlu	132		
405	CTCGCCAGGAGTGTGGGCTTGCTTCCCTGGGGGATTTTCATGGAA	454		
132	gslGlyLysAspTrpGlnGlnThrGlnLysIleTyrAsnCysHisValLeu	149		
455	AGGACCCAACTGGGACAGGACGACCAACCATTTCCAAATTCACAGTGGTTG	504		
149	eunAsnSerLysGlyAlaValAlaThrTrpArgLysThrHisLeuCys	165		
505	TGGACAAACAGGCGACATAGTATCGGTGCCCAAGCTCACCTGT	554		
166	AspValIleGluIleProGlnGlnGlyProMetCysGluSerAsnSerThr	182		
555	GACGTAGACTTGCAGAAATGAGTGTCACTACGAGAGAGCAAGTTCCACCT	604		
182	ProGlyProSerLeuGlnSerProValSerThrProAlaGlyLysIleG	199		
605	CCCCGGACACAGACTTATCCGCCCAATCTCTCCACAGAGAAAGTTG	654		
199	LysLeuAlaValCysTrpAspMetArgPheProGluLeuSerLeuAlaLeu	215		
655	GCTGTGGGGGTGTATCACACCTCCGCTTCCACGAAATTTCTCTGGGCTTG	704		
216	AlaGlnAlaGlyAlaGlnIleLeuThrThrProSerThrIlePheGlySer	232		
705	GCCCAACAGGAGCAACTTCTACCTTACCTCTTCCCTTCAACCTCCAC	754		
232	ethrGlyProAlaHisTrpGluValIleLeuLeuArgAlaArgAlaIleGlu	249		
755	TACTGTGTGGCACTTGGAGGTTGCTGAGAGCCCGGTCATAGAA	804		
249	hrgLysCysTrpValValAlaAlaAlaGlnCysGlyArgHisHisGlyLys	265		
805	CCCAAGCTCTAGTATGTTGCACGCGCACAGACAGCAACAAATGAGAA	854		
266	ArgAlaSerTrpGlyLysSerMetValValAspProThrGlyThrVal	282		
855	AGGAGCTCTTATGGCAAGCTATGATGTGTATGACCCGAGGGGGCTGGCAT	904		

THIS PAGE BLANK (USPTO)

Quality: 1743.00 Length: 326
Ratio: 5.347 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-357-675c-21 x AA246101 ..

Align seg 1/1 to: AA246101 from: 1 to: 1416

```

2 LeuGIyPheIIeThrArgProHIsArgPheIleuSerLeuIeuCysPr 18
|||||
111 CGGGGCTTCATCACCAGGCGCTCTCAGACGATTCCTGCTCTGTC 160
18 oGIyLeuArGIleProGIleuSerValIeuCysAlaGIProArGIPro 35
|||||
161 TCGACTCCGAGATCCTCAACTCTCAGTACTTGTGCTCAGCCAGCCCA 210
35 rGIAlMeAlAlIeSerSerSerSerCysGIuLeuProLeuValAlAl 51
|||||
211 GAGCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 260
52 CysGIuValAlThrSerThrProAspIysGIuIAsnPhelYsthrCysAl 68
|||||
261 TGCACGATGACATCGACGCCAGACAAGCAACTTTAAACATGTGC 310
68 aGIuLeuValAlArGIuAlAlAlArGIuAlIaCysLeuAlAlPheLeuP 85
|||||
311 TGAGCGTGTTCAGAGAGCGTGCAGACTGGGCTCCCTGGCTTCTCTCC 360
85 roGIuAlAlPheAsPheIIeAlAlArGAsPProAlAGIuThrIleuHIsleu 101
|||||
361 CGAGGACATTTGATTCATTCAGCGGACCCCTGACAGACGCTACACCTG 410
102 SerGIuProLeuGIyGIyLysLeuIeuGIuGIuTyrThrGIuLeuAlAlAr 118
|||||
411 TTGTGAACCACTGGGTGGGAACTTTGGAGAAATACACCCAGCTTCCAG 460
118 gGIuCysGIyLeuTrIleuSerLeuGIyGIyPheHIsGIuArGIyGIuAl 135
|||||
461 GGAATGTGAGACTCTGGCTCTCTCTGGGTGTTTCATGAGCGGCGCAAG 510
135 sPTIrGIuGIuThrGIuIleTyrAsnCysHIsValIleuIeuAsnSer 151
|||||
511 ACTGGAGCAGACTCAGAAATCTCAATTTGACGCTGCTGTAACACAC 560
152 LysGIyAlAlAlValAlAlArThrTyrArGIyThrHIsIleuCysAspValGI 168
|||||
561 AAAGGGGCGATGAGTGGCCACTTACAGGAAGACACATCTGTGACGTAGA 610
168 uIleProGIyGIuGIyProMeIcYsGIuSerAsnSerThrMetProGIyP 185
|||||
611 GATTCACAGGCGAGGGGCTATGTGTGAAGCAACTTACATGCGCTGGGCG 660
185 roSerLeuGIuSerProValSerThrProAlAGIyLysIleGIyLeuAlAl 201
|||||
661 CCAGTCTTGAGTCACTCTGACACACAGCCAGGCGCAAGATTTGTCTAAGCT 710
202 ValGIyTyrAspMetArGPheProGIuIleuSerLeuAlAlAlGIuAl 218
|||||
711 GTCTGTATGACATGCGGTTCCCTGACACTCTCTGCGACTTGGCTCAAGC 760
218 aGIyAlaGIuIleLeuThrTyrProSerAlAlPheGIySerIleThrGIyP 235
|||||
761 TCGACACAGAGATACCTTACCTTACCTTGGATCCATTTACAGGCGC 810
235 roAlAlHIsTrIleuValIleuLeuAlAlAlAlAlAlIleGIuThrGIuCys 251
|||||
811 CAGCCCACTGAGGAGGTGTCTGCGGCGCCCTGCTATCCAAACCCAGTGC 860
252 TyrValAlAlAlAlAlAlAGIuIleCysGIyArGIHIsGIuIleArGIAlase 268
|||||
861 TATGTATGTGCGACGACGACAGTGTGAGCGCCACCATGAGAGAGACGAC 910

```

```

268 rTyrGIyHIsSerMetValAlAsPProTrIPGIyThrValAlAlArGC 285
|||||
911 TTATGGCCACACAGCATGTGTGTAGACCCCTGGGGAACAGATGTGGCCGCT 960
285 ySSerGIuGIyProGIyLeuCysLeuAlAlArGIleAsPLeuSnTyrIleu 301
|||||
961 GCTGTGAGGGGCGAGGCGCTCTGCTTGGCCGAAATGACCTCAACTATCTG 1010
302 ArGIuIleuArGIHIsIleuProValIPheGIuHIsArGIArGIProAsPle 318
|||||
1011 CGACAGTTGGCGCCGACACCTGCTGTGTTCAGACACCGAGGCGCTGACT 1060
318 uTyrGIyAsnLeuGIyHIsProIeuSer 327
|||||
1061 CTATGCAATCTGGGTCAACCCACTGTCT 1088

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAF16257
seq_documentation_block:
ID AAF16257 standard; cDNA; 1382 BP.
AC
XX AAF16257;
DF
XX 13-MAR-2001 (first entry)
DE
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
KW
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulnery; gastrointestinal; nephrotropic; antinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
OS
XX Homo sapiens.
PN
XX WO200055174-A1.
PD
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05988.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE-) ROSEN C A.
XX
XX Rosen CA, Ruben SM.
XX
XX MPI: 2000-587513/55.
XX P-PSDB: AAB57054.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
XX disorders such as prostate cancer -
XX
XX Claim 1: Page 1124: 2338bp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytosolic,
XX cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,
XX nephrotropic, antinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF15506 to AAF16514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.
XX

```


50 Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

alignment_scores:

Quality: 1718.00 Length: 327
Ratio: 5.270 Gaps: 0
Percent Similarity: 99.694 Percent Identity: 99.694

alignment_block:

US-09-357-675c-21 x AAF16257 ..

Align seg 1/1 to: AAF16257 from: 1 to: 1382

```

1 MetLeuGlyPheIleThrArgProProHisAspPheLeuSerLeuLeu 17
|||||
77 ATGCTGGGCTTCATCCACAGGCTCTCCACAGATTCCTGCTCCCTGTG 126
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArg 34
|||||
127 TCCTGGAGCTCCGGATACCTCACTCTCAGTACTTGTCTGCTCAGCCAGC 176
34 roArgAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuValAla 50
|||||
177 CCAAGAGCATGGCTATGCTCTCTCTCTGCGAAGTGCCTGCTGCTGCT 226
51 ValCysGlnValIleThrSerThrProAspGlnGlnIleAsnPhelyThrCy 67
|||||
227 GTGTGGCAGGTAACATCATGACGCGACAGACAGCAGCAAGCACTTAAACATG 276
67 sAlaGlnLeuValArgGlnAlaAlaArgLeuGlyAlaCysLeuAlaPheI 84
|||||
277 TGCTGAGCTGTTGAGAGAGCTGCCAGACATGGGCTGCTGCTGCTTTC 326
84 euProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThrLeuHis 100
|||||
327 TGCTGAGGATTTGACTTCATTCACAGGAGCCTGCAAGAGAGCTACAC 376
101 LeuSerIleProLeuGlnGlyIleLeuLeuGlnGlnIleThrGlnLeuAl 117
|||||
377 CTGTGTGAACCATCTGGGTGGAACTTTTGGAAATACACCCAGCTTGC 426
117 aArgGlyCysGlyLeuThrPheLeuSerLeuGlyIlePheHisGlnArgGly 134
|||||
427 CAGGGAATGTGAGACTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 476
134 InAspTrpGlnGlnThrGlnIleGlyIleGlyIleGlyIleGlyIle 150
|||||
477 AAGACTGGAGACAGACATCAGAAATTAACAATTCACAGCTGCTGCTGAC 526
151 SerIleGlyAlaValAlaIleThrIleArgIleThrHisLeuGlyAspVal 167
|||||
527 AGCAAAAGGCGAGTACAGGCTTACAGGAAGACATCTGTGTGAGCT 576
167 IglIleProGlyIleGlnIleProMetCysGlnSerAsnSerThrMetProG 184
|||||
577 AGACATTCAGAGGAGGAGGCT.ATGTGTGAAGCAACTCTACCTATGCTG 625
184 LyProSerLeuGlnSerProValSerThrProAlaGlyIleGlyIleGly 200
|||||
626 GCCCAATCTTGAAGTCACTGTCAGACACAGCAGCAGCAATGAGTGTGA 675
201 AlaValIleGlyIleAspMetArgPheProGlnLeuSerLeuAlaLeuAla 217
|||||
676 GCTGTCTGATGACATGCGGCTTCCCTGAACTCTCTGAGGATGAGTCA 725
217 nAlaGlyAlaGlnIleLeuThrIleThrProSerAlaPheGlySerIleThr 234
|||||
726 AGCTGGAGCAGAGTACTTACTTACCTTCACTTGTGAGATCAATTAAG 775
234 LyProAlaHisTrpGlnValLeuLeuArgAlaAlaGlnIleGlnIleThr 250
|||||
776 GCCCAGGCGCACTGGAGGTGTGCTGGGCGGCGGCTGATGCAACACGAG 825

```

```

251 CysTrpValValAlaAlaAlaGlnCysGlyIleArgHisIleGlnIleArgAl 267
|||||
826 TGCTATGATGAGCGACAGACAGCTGTGAGACCCACCACTGAAAGAGAGC 875
267 aserTrpGlyHisSerMetValValAspProTrpGlyThrValValAlaAla 284
|||||
876 AAGTTATGGCCACAGCATGCTGTAGACCCCTGGGAAACATGTGTTGGGCC 925
284 rGlySerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTrp 300
|||||
926 GCTCTCTGAGGGGCGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 975
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
|||||
976 CTGGAGACAGTGGCGCGACACCTGCTGTGTTCAGACACCGAGAGCTGA 1025
1026 CCTCTATGGCAATCTGGGTGACCCACTGTCT 1056
317 PleuTrpGlyAsnLeuGlnHisProLeuSer 327
|||||
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AA30398
seq_documentation_block:
ID AAX30398 standard; DNA; 1203 BP.
XX
AC AAX30398;
XX
DT 14-MAY-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS Homo sapiens.
XX
PN MO9907891-A1.
XX
PD 18-FEB-1999.
XX
PF 04-AUG-1998; 98MO-US16235.
XX
PR 19-AUG-1997; 97US-0056732.
PR 05-AUG-1997; 97US-0054798.
PR 05-AUG-1997; 97US-0054803.
PR 05-AUG-1997; 97US-0054804.
PR 05-AUG-1997; 97US-0054806.
PR 05-AUG-1997; 97US-0054807.
PR 05-AUG-1997; 97US-0054808.
PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055386.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056365.
PR 19-AUG-1997; 97US-0056366.
PR 19-AUG-1997; 97US-0056557.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Ferlie AM, Greene JM, Janat F, Ni J,
Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;

```

XX WPI: 1999-167452/14.
 DR P-PSDB: AAV10877.
 XX New isolated human genes encoding secreted polypeptides - useful for
 PT diagnosis and treatment of pathological diseases
 XX
 PS Claim 3; Page 265-266; 331pp: English.

XX The specification describes secreted proteins and their corresponding
 CC polynucleotides which are useful for preventing, treating or ameliorating
 CC medical conditions, e.g. by protein or gene therapy. Pathological
 CC conditions can also be diagnosed by determining the amount of the
 CC secreted polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the products, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC diseases, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,
 CC prostate diseases, asthma, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
 CC thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.

SQ Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

alignment_scores:
 Quality: 1067.00 Length: 207
 Ratio: 5.205 Gaps: 0
 Percent Similarity: 99.034 Percent Identity: 99.034

alignment_block:
 US-09-357-675c-21 x AAX30398 ..

Align seg 1/1 to: AAX30398 from: 1 to: 1203

121 GYLLEUTRPLEUSERTLEUGLYGYPHEHISGLUARGLYSLNAPRPGI 137
 ||||||||||||||||||||||||||||||||||||||||||||||||
 3 GGACTGTGGCTGCTTGGGTGGTTCCATGACGCGTGCCCAACATGGGA 52
 ||||||||||||||||||||||||||||||||||||||||||||||||
 137 uGlnTrGlnLysIleTyrAsnCySHisValLeuLeuAsnSerLysGly 154
 ||||||||||||||||||||||||||||||||||||||||||||||||
 53 GCAGACTTCAGAAATCTACAAATGTCTACGCGTGCAGAAAG. AAGGGG 101
 ||||||||||||||||||||||||||||||||||||||||||||||||
 154 lAValAlaIaThrTyrArgLysThHisLeuCySAspValGluIlePro 170
 ||||||||||||||||||||||||||||||||||||||||||||||||
 102 CAGTACTGGCCACTTACAGAGAACACATCTGTGTGACGTAGAGATCCCA 151
 ||||||||||||||||||||||||||||||||||||||||||||||||
 171 GLYGLNGLYPROMETCYSGIUSERASERTHMETPROGLYPROSERILE 187
 ||||||||||||||||||||||||||||||||||||||||||||||||
 152 GGGCAGAGGGCT. ATGTGTGAAGCAACTTACCATGCTGGCCCACTCT 200
 ||||||||||||||||||||||||||||||||||||||||||||||||
 187 uGUserProValSerThrProLagLysIleGlyLeuAlaValCysT 204
 ||||||||||||||||||||||||||||||||||||||||||||||||
 201 TGTGTCACCTGTCAGACACACACAGCAAGATTGGTCTAGCTGCTGCT 250
 ||||||||||||||||||||||||||||||||||||||||||||||||
 204 YrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
 ||||||||||||||||||||||||||||||||||||||||||||||||
 251 ATGACATGGGGTTCCTGCAACTCTCTGTGCAATTGGCTCAAGCTGGAGCA 300
 ||||||||||||||||||||||||||||||||||||||||||||||||
 221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaH 237
 ||||||||||||||||||||||||||||||||||||||||||||||||
 301 GAATATCTTACCTATCTTCAGCTTTGGATCCATTACACAGCCACAGCCA 350
 ||||||||||||||||||||||||||||||||||||||||||||||||
 237 STRPGLIValLeuLeuArgAlaArgAlaIleGluTrnGlnCysTyrValY 254
 ||||||||||||||||||||||||||||||||||||||||||||||||
 351 CTGGAGAGTTCCTGCGGGCCGCTGTATGCAACCACTGCTATGTAG 400
 ||||||||||||||||||||||||||||||||||||||||||||||||
 254 aAlaAlaAlaGlnCysGlyArgHisGluLysArgAlaSerTyrGly 270

||||||||||||||||||||||||||||||||||||||||||||||
 401 TGGCAGCAGCAGCAGTGTGGACGCCACCATGAGAAAGACCACTTATGGC 450
 ||||||||||||||||||||||||||||||||||||||||||||||||
 271 HisSerMetValValAspProTrpGlyThrValValAlaArgCysSergI 287
 ||||||||||||||||||||||||||||||||||||||||||||||||
 451 CACAGCATGTGTGTAGACCCCTGGGGACAGTGTGTGCTGCTCTCTGA 500
 ||||||||||||||||||||||||||||||||||||||||||||||||
 287 uGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnL 304
 ||||||||||||||||||||||||||||||||||||||||||||||||
 501 GGGGCCAGGCGCTCTGCTTGGCCGAATAGACCTCACTATCTCGACAGT 550
 ||||||||||||||||||||||||||||||||||||||||||||||||
 304 euArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320
 ||||||||||||||||||||||||||||||||||||||||||||||||
 551 TGGCGGACACCTGCTGTGTTCACACACCGCAGCGCTGACCTTATGGC 600
 ||||||||||||||||||||||||||||||||||||||||||||||||
 321 AsnLeuGlyHisProLeuSer 327
 ||||||||||||||||||||||||||||||||||||||||||||||||
 601 AATCTGGGTACCCACTGTCT 621

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD25458

seq_documentation_block:

ID AAD25458 standard; cDNA; 1214 BP.

XX AAD25458;

XX 26-MAR-2002 (first entry)

XX Xenopus laevis Nt11 cDNA.

XX Frog; Nt11; cytosolic; neuroprotective; cellular pathway; therapy;

KW apoptosis; proliferative disorder; degenerative disease; ss.

XX Xenopus laevis.

XX Key Location/Qualifiers

FT CDS 180..1046

FT /tag= a

XX /product= "Frog Nt11 protein"

XX WO200187958-A2.

XX 22-NOV-2001.

XX 15-MAY-2001; 2001WO-US15664.

XX 16-MAY-2000; 2000US-204713P.

XX (UIJE-) UNIV JEFFERSON THOMAS.

XX Croce C, Brenner C, Pekarski Y;

XX WPI: 2002-082984/11.

XX P-PSDB: AAE15792.

XX Isolated cDNA encoding human, mouse, frog and yeast Nt2 proteins,

XX useful to find molecules that mimic or antagonize Fhit interaction for

XX the treatment of proliferative or degenerative diseases -

XX Claim 21; Page 60; 61pp: English.

XX The invention relates to isolated nucleic acids comprising a fully

XX defined cDNA nucleotide sequence encoding human, Xenopus laevis and

XX mouse Nt2 proteins. Nt1 and Fhit proteins are encoded as fusion

XX proteins in invertebrates and as separate polypeptides in vertebrates.

XX Nt1 and Fhit interact physically and functions in same cellular pathways.

XX Molecules which bind Nt2 and mimic or antagonize Fhit interaction are

XX used to treat diseases in which activity of Nt2 protein is altered in

XX a mammal. Fhit mimics induce apoptosis and are particularly useful to

XX treat proliferative disorders, whilst Fhit antagonists promote cell

XX proliferation and are particularly useful to treat degenerative disease.

XX The present sequence is frog Nt11 cDNA.

Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

alignment_scores:

Quality: 968.50 Length: 307
Ratio: 3.905 Gaps: 1
Percent Similarity: 80.782 Percent Identity: 57.980

alignment_block:

US-09-357-675c-21 x AAD25458 ..

Align seg 1/1 to: AAD25458 from: 1 to: 1214

```

16 LeucylProglyLeuArgIleProGluLeuSerValLeuCysAlaGlnPr 32
   |||||||.....:|||||.....
144 CTGTGTCGGGATATAGGGGTTAGAAAGCTGTAGTCAATG..... 182
   |||||||.....:|||||.....
32 CArgProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuV 49
   |||||||.....:|||||.....
183 .....GCTGGTCCCAACAGCCCTGA 204
   |||||||.....:|||||.....
49 AlaAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhel 65
   |||||||.....:|||||.....
205 TTGCGGTGCGCAGATGACTTCAACCTCTGATAGAGAAAGAAATTGCGC 254
   |||||||.....:|||||.....
66 ThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAl 82
   |||||||.....:|||||.....
255 ACGGTTCGGCGGTGATCCGGAGGCTCCGGGCGTCCGCTTCATGCT 304
   |||||||.....:|||||.....
82 alphaLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThr 99
   |||||||.....:|||||.....
305 GTTTCCTCCGGAAGCCTTTGACTATATCGGGGAGCATTAGAGAGAGCC 354
   |||||||.....:|||||.....
99 euHisLeuSerGluProLeuGlyLysLeuLeuGluGluThrGln 115
   |||||||.....:|||||.....
355 TGAGTCTGGCTGAGTCTCTACATGGGGACACCATTCACCGTTACACCAA 404
   |||||||.....:|||||.....
116 LeuAlaArgGlyCysGlyLeuThrLeuSerLeuGlyPheHisGluArg 132
   |||||||.....:|||||.....
405 CTCGCCAGGAGGTGGGCTGGCTTCCTCGGGGATTCATGAGAAA 454
   |||||||.....:|||||.....
132 gGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuL 149
   |||||||.....:|||||.....
455 AGGACCCCAACTGGGACACGACGCAACGATTTCCATTTCTCACGGGTG 504
   |||||||.....:|||||.....
149 euAsnSerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeu 165
   |||||||.....:|||||.....
505 TGGACACACAGGACACATAGTACGTACCGCAAGGCTCACCGTTT 554
   |||||||.....:|||||.....
166 AspValGluIleProGlyGlnGlyPrometCysGluSerAsnSerThrMe 182
   |||||||.....:|||||.....
555 GACGTAGCTTGCAGAAATGAGTGTCACTCAGAGAGCACTTCACCCCT 604
   |||||||.....:|||||.....
182 TrpGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 199
   |||||||.....:|||||.....
605 CCCGGGACAGACGCTTATTCGCCCATCACTTCTCCACAGAAAGATTG 654
   |||||||.....:|||||.....
199 LysLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeu 215
   |||||||.....:|||||.....
655 GCGTGGGGGTGTGTACGACCTCCGCTCCCAAGATTCTCCTGGCTGTG 704
   |||||||.....:|||||.....
216 AlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerI 232
   |||||||.....:|||||.....
705 GCCCAACAGAGGACGACACTTCTCACTTACCTTCCTGCTCCATCCCTC 754
   |||||||.....:|||||.....
232 eThrGlyProAlaHisTrpGluValIleLeuAlaArgAlaArgAlaIle 249
   |||||||.....:|||||.....
755 TACGTGTGTGACATGTGGAGGTGTTCGTAGAGCCGCTGCCATAGAAA 804
   |||||||.....:|||||.....
249 hrGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGluLys 265
   |||||||.....:|||||.....
805 CCCAGTGTCTAGTAGTGTGACAGCGCACAGACAGACAGACAAATGAGAAG 854

```

```

266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVa 282
   |||||||.....:|||||.....
855 AGGACGTCTATGTGTACGCGTATGTGTAGACCCGTGGGGCTGTGTCAT 904
   |||||||.....:|||||.....
282 LAlaArgCysSerGluGlyProGlyLeuCysLysLeuAlaArgIleAspLeu 299
   |||||||.....:|||||.....
905 TGGCCAAATGCCAGAGACAGAAATATGTTATGCTGATGACATTCATTC 954
   |||||||.....:|||||.....
299 snTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArg 315
   |||||||.....:|||||.....
955 CCTACATGGAGCGGTGTAGCGGACATGCGGTGTGGAGCACCAGG 1004
   |||||||.....:|||||.....
316 ProAspLeuTyrGlyAsnLeu 322
   |||||||.....:|||||.....
1005 ACTGATCTGTATGGGAAATATC 1025
   |||||||.....:|||||.....
seq_name: /sids1/gcgdata/hold-geneseg/geneseqn-emb1/NA2001B.DAT:ABL12225
seq_documentation_block:
ID ABL12225 standard; cDNA; 1495 BP.
XX
XX ABL12225;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001MO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-P-SDB; ABB68122.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX
XX interactions -
XX
XX Claim 1; SEQ ID NO 31157; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX
XX useful in developmental biology and in elucidating cell signalling and
XX
XX cell-cell interactions in higher eukaryotes for the development of
XX
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX
XX sequences (AB101840-AB116175) and the encoded proteins
XX
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX
XX specification, but was obtained in electronic format directly from WIPO
XX
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;

```

alignment_scores:

Quality: 710.50 Length: 294
Ratio: 3.259 Gaps: 3
Percent Similarity: 74.150 Percent Identity: 48.639


```

1183 GGATCAAAAGTGGACCATTTGCTGTGGTGCAGATGCCCTCAGCAGGAC 1232
60 ysgInGlnAsnPhelysthrcysAlaGluLeuValArgGluAlaAlaArg 76
1233 AGCGCGCTATCTTTAGTCAAGTGTAGAGCTAGTGTAGTGGCCAGTCA 1282
77 LeuGlyAlaCysLeuAlaPhelProGluAlaPhaAspPheIleAla 93
1283 CAGAACGCTGCATGCTTTCTTCTGCTGAGTGTGCTGACTTTGTGGCGA 1332
93 gaspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeu 110
1333 GAGCGGACCCAAACATATGAGCTCTCCAGCGGCTTGACGCGGAGTTAA 1382
110 euGluGluThrThrGlnLeuAlaArgGluCysGlyLeuThrPleuSerLeu 126
1383 TGGCGCAGTACCGGAAATTGGCAAGTGCACAAATTTGATTTCCCTG 1432
127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnIleTy 143
1433 GGTGGCTGCAGCAGCGGAACT.....CAAAAAATCTT 1467
143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaThrTyra 160
1468 CAACGCTCATGTTTGTCTCAGCAGAAAGGAGGAGTACGACGAGTACA 1517
160 rGlyThrHisLeuCysAspValGluIleProGlyGlnGlyPrometCys 176
1518 GAAAGCTGCACATGTTGATGTACGACTAAA...GAGGTCCGCTACGC 1564
177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProAlaSerh 193
1565 GATCTGAGTACAGTTAGCGCGGATCTCTGAGCGCCCATGAGCAGC 1614
193 rProAlaGlyLysIleGlyLeuAlaValCysTyraSpmetArgPhePro 210
1615 TCCAGTGGCCAGATAGGGCTTCAGATTTGCTACGACGCTGCTTTGGCG 1664
210 LuLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyro 226
1665 ACCCGCGCTGCTGCTCAGGAAAGCTGGGCCAATTTGTTAAATATCCCA 1714
227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeu 243
1715 TCCGCAATTCATATACGCAACCGGTAAGCGCACTGGAAATCCTCTGCG 1764
243 gAlaArgAlaIleGluThrGlnCysTyraValAlaAlaAlaGlnCysG 260
1765 GGCAGAGCCATAGAGACTCAATGCTTTGGTGGCTGCGCGCTCAATAG 1814
260 LysrGHisHisGluLysArgAlaSerTyroLysHisSerMetValAla 276
1815 GTTGGCAGCAGCAGACGACAGATGTTGGGCGCACAGCATGATCGTTAG 1864
277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeu 292
1865 CCTGGGGAACGTAAGCTGCTGACTCAGCGAGCAGAGCTGATATAG 1914
292 sLeuAlaArgIleAspLeuAsnTrpLeuArgGlnLeuArgArgHisLeu 309
1915 CAGCGCGCAGAGTGCAGCTTCCCTGCTCAATCTCTGTATCAGACCATG 1964
309 roValPheGlnHisArgArgProAspLeuTyro 319
1965 CCTGCTTCGAACTCGTCAAAACGACATCATC 1996
seq_name: /SID1/sgdata/hold-geneseq/geneseg-emb1/NA2001B.DAT:ABL20222
seq_documentation_block:
ID ABL20222 standard; DNA; 5692 BP.
XX
AC ABL20222;

```

```

XX
DE 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12139; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB1616175) and the encoded proteins
CC sequences (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WPI
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;

```

```

alignment_scores:
  Quality: 710.50      Length: 294
  Ratio: 3.259         Gaps: 3
  Percent Similarity: 74.150   Percent Identity: 48.639

```

```
alignment_block:
US-09-357-675C-21 x ABL20222 ..

```

```
Align seg 1/1 to: ABL20222 from: 1 to: 5692

```

```

27 ValLeuCysAlaGlnProArgProAlaMetAlaIleSerSerse 43
   |||:||||: ||: ||| |||:||||: ||:
2792 GTGATGCCATTCATCAGCAGACTACGAGTATCGCTCCAAAAGAGGA 2841
43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAsp 60
   : : : : : : : : : : : : : : : : : :
2842 GGATCAAAAGTGCAGCAATTTGCTGTGGTGCAGATGCCCTTACACGACA 2891
60 ysgInGlnAsnPhelysthrcysAlaGluLeuValArgGluAlaAlaArg 76
   |||:||||: |||: ||| |||:||||: ||:
2892 AGCGCGCTATCTTTAGTCAAGTGTAGAGCTAGTGTAGGCGCAGTCA 2941
77 LeuGlyAlaCysLeuAlaPhelProGluAlaPhaAspPheIleAla 93
   : : : : : : : : : : : : : : : : : :
2942 CAGAACGCTGCATGCTTTCTTCTGCTGAGTGTGCTGACTTTGTGGCGA 2991
93 gaspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeu 110

```



```

XX 15-MAY-2001; 2001MO-US15664.
PE
XX
XX 16-MAY-2000; 2000US-204713P.
PR
XX
XX (UYJE-) UNIV JEFFERSON THOMAS.
PA
XX
PI Croce C, Brenner C, Pekarski Y;
XX
XX WPI; 2002-082984/11.
DR
XX P-PSDB; AAE15790.
XX
PT Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,
PT useful to find molecules that mimic or antagonise Fhit interaction for
PT the treatment of proliferative or degenerative diseases
PT
XX
PS Claim 3; Page 59; 61pp; English.
XX
CC The invention relates to isolated nucleic acids comprising a fully
CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and
CC mouse Nit2 proteins. Nit and Fhit proteins are encoded as fusion
CC proteins in invertebrates and as separate polypeptides in vertebrates.
CC Nit and Fhit interact physically and functions in same cellular pathways
CC Molecules which bind Nit2 and mimic or antagonise Fhit interaction are
CC used to treat diseases in which activity of Nit2 protein is altered in
CC a mammal. Fhit mimics induce apoptosis and are particularly useful to
CC treat proliferative disorders, whilst Fhit antagonists promote cell
CC proliferation and are particularly useful to treat degenerative disease.
CC The present sequence is human Nit2 cDNA.
XX
XX Sequence 1359 BP; 388 A; 284 C; 312 G; 374 T; 1 other;
XX
alignment_scores:
Quality: 485.50 Length: 286
Ratio: 2.542 Gaps: 6
Percent Similarity: 66.783 Percent Identity: 36.713
XX
alignment_block:
US-09-357-675C-21 x AAD25456 ..
XX
Align seg 1/1 to: AAD25456 from: 1 to: 1359
XX
49 VALAAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhely 65
:::|||||::: |||||::: :: |||||:::|
37 TTGGCCCTTCATCCACGCTTCAGATTCTTCATCAAAATCAGATACGTCAC 86
:::|||||:::|
65 sThrcysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuA 82
:::|||||:::|
87 TCGCCCTTGATAGCTTCATCCGGAGGCGACGACCAACGAGACCAAAATAG 136
:::|||||:::|
82 LaPhLeuProGluAlaPhaLeaspheLeAlaArgAspProAlaGluThr 98
:: |||||:::|
137 TTCTCTTGCCGGAATGCTTT ..... 156
:::|||||:::|
99 LeuHISLeuSerGluProLeuGlyGlyLysLeuLeuGlnGluTyr.... 113
:::|||||:::|
157 .....AATCTCATATGAGAGCGAAATATTTCTGTAATGAGAGA 197
:::|||||:::|
114 .....ThGlnLeuAlaArgGluC 120
:::|||||:::|
198 GAAATTCCTGTGTAATCCACAGACAGACGCTTTCGAGATACCAAGGAAAT 247
:::|||||:::|
120 YAGGLeuThrLeuSerLeuGlyGlyPheHISGluArgGlyGlnAspTyr 136
:::|||||:::|
248 GAGCATATATCTCATTTGGAGGCTCTATCCCTGAA..... 282
:::|||||:::|
137 GluGlnThrGlnLysIleTyrAsnCysHISValLeuLeuAsnSerTysG1 153
|||:::|||||
283 GAGGATGCTGGGAATTTATATACACCGCTGCTGTTTGGCGCTGATGG 332
|||:::|||||
153 ValAValAlaIaThrTyrArgGlyThrHISLeuCysAspValGluIleP 170
|:::||||| ||||| |||||

```



```

129 .....PheHISgluArgglyInAspTrpIugInThrGlu 141
XX      ||| |||
26886 TGGCGCGTCAAAAGCTGGAGCGGGT.....A 26858
141 ysIeIyrnscYsHISValleuLeuAsnSerIysglYalValAla 157
XX      ||| |||
26857 AAGTGATGATACGCTGTGTGTACGAGCGGCGGCAAGACGGGG 26808
158 ThrTYArArgIysThrHIScysAspValGluIleProGlyGlnGly 174
XX      ||| |||
26807 CAGTACCACAAAATGCACCTC.....TTCGGTTTCCGGTTGGGGCA 26764
174 oMetCysgluSerAsnSerThrMetProGlyProSerleuGluSerPro 191
XX      ||| |||
26763 ACGCTATGCCGAAGCGATACATCCGCGGGGGAATGTCGCCGAC 26714
191 aISerThrProAlaGlyIysIleGlyLeuAlaValcysTyraSPMetArg 207
XX      ||| |||
26713 TCTCGCAAGAGCGGTGCGGGGCGGCGGCAATTGTGTACGATGTCGCG 26664
208 PheProGluLeuSerleuAlaLeuAlaGluAlaGlyAlaGluIleuTh 224
XX      ||| |||
26663 TTTCGCCGA.....TTTTCGACGCCGAGTTGCCGTTGACGTATTGAT 26620
224 rTYrProSerAlaPheGlySerIleThrGlyProAlaHISTrpGluVal 241
XX      ||| |||
26619 GCTGCCCGCTGCGTTTACGACACGACGAGGCAAGCGCATTTGGAGCTCG 26570
241 euLeuArgAlaArgAlaIleGluThrGlnCysTyValAlaAlaAla 257
XX      ||| |||
26569 TCTGTGCGCGCGGTGCGGCAAAACCAATGTACCTGTGCGGGCGGA 26520
258 GlnCysGlyArgHISgluIysArgAlaSerTYrGlyHISserMetVal 274
XX      ||| |||
26519 CAGGGGGGTTCGACAAACGACGCGCACCTTCGACACGACATGAT 26470
274 IValAspProTrpGlyThrValAlaAlaArgCysSerGluGlyProGly 291
XX      ||| |||
26469 TCTCGATCCGTGGGCGGCGATGTTGGACGTAATGCCGAGGCGAAGGCG 26420
291 euCysleuAlaArgIleAspLeuAsnTYrleuArgGlnleuArgAlaG 307
XX      ||| |||
26419 TTGTTACGCGACATGATGCCAACCCTGAAACGCTCGCAACCGC 26370
308 LeuProValPheGlnHISArg 314
XX      ||| |||
26369 CTGCCCGCTTGAAATACCGG 26349
seq_name: /SIDSL/gcgsdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAF21607
seq_documentation_block:
ID AAF21607 standard; DNA; 349980 BP.
XX
AC AAF21607;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PE 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.

```

```

PR 28-FEB-2000; 2000GB-0004695.
XX
XX (CHTR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V,
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappelli R;
PI Frazer CM, Grandi G;
XX
XX WPI: 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
PS Claim 7; Appendix A; 692pp; English.
XX
XX
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
XX Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;
SQ

```

```

alignment_scores:
Quality: 404.00 Length: 307
Ratio: 2.138 Gaps: 8
Percent Similarity: 61.564 Percent Identity: 35.179

```

alignment_block:

Align seg 1/1 to: AAF21607 from: 1 to: 349980

```

16 LeuCysProGlyLeuArgIleProGlnLeuSerValleuCysAlaGlnPr 32
||| ||| ||| ||| ||| |||
153960 TTGTGTGMAA.....CGTTTTCGCAAAAGTACCGTCATGCTACCGTATCG 154003
32 oArg.....ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 47
||| ||| ||| ||| ||| |||
154004 GCGGAATATGCGCGCAAGTCCGCGCAGAAAGAGAAATGACAAAA 154053
47 roLeuValAlaValcysGlnValThrSerThrProAspIysGlnGlnAsn 63
||| ||| ||| ||| ||| |||
154054 TCAGAGTTCCGCGCGCGAGATGCTGTGCGGCGTCCGCGGAACCAAC 154103
64 PheIysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaC 80
||| ||| ||| ||| ||| |||
154104 GTCGCCCGCATGAACGCTGTGCGCACGCGCGGACAGAGTCCGGA 154153
80 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 97
||| ||| ||| ||| ||| |||
154154 TTGGGTGCTGTCGCCGAATATTGGTGTCTGATGGGCGCAACGATACGG 154203
97 IuThrleuHISleuSerGluProLeu...GlyGlyIysleuLeuGluGlu 112

```

```

154204 ACAAACTCGCGCTTGGCGGCGGACGCTTCAACAGCGCA 154253
      :: ||| |||::||| ||| |||::|||
113 TyrThrInleuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
      ::|||::|||::|||::|||::|||::|||
154254 TTGAGCGAAGCGGCGGAAAGATGCGGTGGTGTG...TTGCGCGGCGAC 154300
      ::|||::|||::|||::|||::|||
129 .....phenisgluArgGlyGlnAspTrpGluInthrGlnL 141
      ||| |||
154301 TGTCCCGCTGCAAGCTCGAGCGGCT.....A 154329
141 ysileTyAsnCysHisValleuLeuAsnSerLysglYalaValaVala 157
      |||::|||::|||::|||::|||::|||
154330 AAGGATGAATACCGCTGTGGGTACGACGCGGCGGTAGACGGGG 154379
158 ThrTyArgLysThrHisLeuCysAspValGluIleProGlyInGlyPr 174
      |||::|||::|||::|||::|||::|||
154380 CTGTACACCAAAATGCACCTC.....TTGCGTTTTCGGTTGGCGG 154423
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
      ::|||::|||::|||::|||::|||
154424 ACGGTATGCCAGACCCGATACCATCCGCGCGGGGATGTCGCCACT 154473
191 alserThrProAlaGlyLysIleGlyLeuAlaValaCysTyAspMetArg 207
      ::|||::|||::|||::|||::|||::|||
154474 TGTGCGGCAAGGCGGTGCGGCGGCGGCGCATTTGTACAGATGCCG 154523
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyGluIleLeuThr 224
      |||::|||::|||::|||::|||::|||
154524 TTTCCTCCAA.....TTTTCGACGCGCATTTGCCGTTTGCATTCAT 154567
224 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 241
      |||::|||::|||::|||::|||::|||
154568 GCTGCCCGCTGCGTTTACGACACGACGCGGCAAGCGCATTTGGGCTG 154617
241 euleuAlaAlaArgAlaIleGluThrGlnCysTyValValaValaVala 257
      |||::|||::|||::|||::|||::|||
154618 TGTGCGCGCGCGGTGCGGCGGCGGCAAAATTTACGTGTGGCGGCA 154667
258 GlnCysLeuArgHisHisGlnLysArgAlaSerTyGlyHisSerMetVa 274
      |||::|||::|||::|||::|||::|||
154668 CAGGCGCGTTTGCACGAAACGCGCGGCGCATTTCCGACACAGCATAT 154717
274 lValAspProTrpGlyThrValValaAlaArgCysSerGluGlyProGly 291
      ::|||::|||::|||::|||::|||::|||
154718 TGTGATTCGCGGCGGCGACGTTTGACGATTTGCCCGGAGGCGAAG 154767
291 euCysLeuAlaArgIleAspLeuAsnTyTrpLeuArgGlnLeuArgHis 307
      |||::|||::|||::|||::|||::|||
154768 TTGTACGCGGACATGATGCCAACCGCTGAAACGCGTCCGCAACCG 154817
308 LeuProValPheGlnHisArg 314
      |||::|||::|||::|||::|||::|||
154818 CTGCCCGCTTGAATACCGG 154838
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-embL/NA2000.DAT:AAA81490
seq_documentation_block:
ID AAA81490 standard; DNA; 1437668 BP.
XX
AC AAA81490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KN Meningococcus B; Menb; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.

```

```

XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99NO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CW, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Rattl G, Scarselli M, Scariato V;
PI Rappoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 866-1272; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins. AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis Menb polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

```

```

alignment_scores:
Quality: 404.00 Length: 307
Ratio: 2.138 Gaps: 8
Percent Similarity: 61.564 Percent Identity: 35.179

```

```
alignment_block:
US-09-357-675C-21 x AAA81490 ..

```

```
Align seg 1/1 to: AAA81490 from: 1 to: 1437668
```

```

16 LeuGlyProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
||||| |||::|||::|||::|||::|||
453960 TTGTGTAAA.....CGGTTTCGCAAAAGTACCCTCATGATGCTATCG 454003
32 oArg.....ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 47
||| |||::|||::|||::|||::|||
454004 GCGGAATATTCGCGCAAGGTGCGGCAAGAAAGAGCAAGAAATGCAAAA 454053
47 rOleuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsn 63
|||||::|||::|||::|||::|||::|||
454054 TCAGAGTTGCCCGCTGCAGATGTGTGCGGCGGTGCGCGGAAACCAAC 454103

```


alignment_block:

US-09-357-675c-21 x AAC71716 ..

Align seg 1/1 to: AAC71716 from: 1 to: 1725

```

23 ProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAla11 39
|||||: : : : : |||
11 CCGCAGGTGTGCTGTCTGC.....AGAGTCATG..... 40
39 eSerSerSerCysGlnLeuProLeuValAlaValCysGlnValThrs 56
: : : : : |||: : : : :
41 .....ACCTCTTCGCTTGCGCTTCATCCAGCTT...CAGATTCTCT 80
56 eThrProAspLysGlnGlnAsnPhelYsthrCysAlaGlnLeuValArg 72
|||: : : : : |||
81 CCATC.....AAATCAGATTAACGTACCTCGCGCTTGATGCTTCATCCGG 124
73 GluAlaAlaArgLeuGlnAlaCysLeuAlaPheLeuProGlnValaPheAs 89
|||||: : : : : |||
125 GAGGCGCAGCAGCAGCAGCAGCAAAATAGTTCTTTCGCGAATGCTT.. 172
89 pheIleAlaArgAspProAlaGlnThrLeuHisLeuSerGlnProLeu 106
: : : : : |||
173 .....AATCTCATATG 185
106 lYgLYstLeuLeuGlnGluTyr..... 113
|||: : : : : |||
186 GAGCGAAATATTTCTCTGAATATGACAGAAAATTCCTGCTGAATCCACA 235
114 .....ThrGlnLeuAlaArgGlnCysGlnLeuTrpLeuSerLeuG1 127
: : : : : |||: : : : : |||
236 CAGAAGCTTCTGAAGTAGACAAAGAAATGACATATCTCTATTGGAG 285
127 yclYphe..HisLuarGlyGln..AspTrpGlnInThrGlnLysIle 142
|||: : : : : |||: : : : :
286 TAACCTCCATCCACACAAAGCTCTATCCCTGAAGAGATGCTGGGAAATRA 335
143 TyrAsnCysHisValLeuLeuAsnSerLysGlnAlaValAlaThr 159
|||||: : : : : |||
336 TATTAACACCTGCTGTGTGTGGGCTGATGAACTTACTCAAAAGTA 385
159 rArgLysThrHisLeuCysAspValGluIleProGlnGlnGlyProMetC 176
|||||: : : : : |||: : : : :
386 TAGAAAGATCCATCTGTGACATGATGCTCTCGAAAAATATACATTTC 435
176 ysgLusSerAsnSerThrMetProGlnProSerLeuGlu..... 188
|||||: : : : : |||
436 AAGAACTAAACATGAGTCCGCGGTGATGATTCTCCACATTGATACT 485
188 ..... 188
486 CGTAGTACCAGATAAGTTGCTCTTAGCAATCTCAGTAGAAGACAAT 535
189 .....SerProValSer..... 192
|||||: : : : : |||
536 CAGGTATTATTCTTTTGTCTCTCTCGATTCTCTCACAATACCTAA 585
192 ..... 192
586 CTGAAGACCATTAGTAGAAGAGCAGAGATCATCAGATCTGGAAG 635
192 ..... 192
636 TTTCGGCTTATTGAGAACTAAGATTGACACAGATTGTCCTTGATT 685
192 ..... 192
686 TGATTGACTTCTCTGTTAGCGCTTCCAGATATACCTATTAGGCTACAG 735
192 ..... 192
736 TTGAGTACCTCCACTAGATATAAGCAATTCAATTGAAATGATTCTTC 785

```

```

193 .....ThrProAla..GlyLys..... 197
|||||
786 ATCTTACTTCGCCGTGATGTAATGATGATCTTTATGAGATGAAGTCCAGT 835
197 ..... 197
836 AGGAATGAGCTGTAAATTATCTGTGCTCAGGTCCGTGTTAATTAT 885
197 ..... 197
886 CCCTGTCAGTGTGTTGATCATATTATGATGAGGATTTCCCTCCAC 935
197 ..... 197
936 ACCATGCTGAGGAGTAACTTTTCATTGTGTCATTTTCTGTTGAA 985
198 .....IleGlyLeuAlaValCysTyrAspMetArgPhePro 209
: : : : : |||: : : : : |||
986 CACCTTACTGACAGATGGGTGCGCATCTGTACGACATGCGTTGCA 1035
210 GluLeuSerLeuAlaLeuAlaGlnAlaGlnIleLeuThrTyrPr 226
|||||: : : : : |||
1036 GACCTTGACAAATCTACGCACAGAGAGGCTGCCACCTGTGTTGATATTC 1085
226 oSerAlaPheGlySerIleThrGlyProAlaHisTrpGlnValLeuLeuA 243
|||||: : : : : |||
1086 AGGAGCTTTTATCTGACCCCTGACACGCCCATTTGGAGTTACTTCAC 1135
243 rGlaAlaArgAlaIleGlnThrGlnCysTyrValAlaAlaAlaGlnCys 259
|||||: : : : : |||
1136 GAGCGCGGCTGTGATATACAGGTGATGTCACAGCCTCTCCGTC 1185
260 GlyArgHisHisGlnLysArgAlaSerTyr.....GlyHisSerMe 273
|||||: : : : : |||
1186 .....CGGATGACAAAGCCTCTATGTTGCCCTGGGGACACACAC 1226
273 tValValAspProTrpGlyThrValValAlaArgCysSerGlnGlyProG 290
|||||: : : : : |||
1227 CGTGGTGAACCTTTGGGGAGGTTCTAGCCAAAGCTGGCAGACAGAGAG 1276
290 lYleuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArg 306
|||||: : : : : |||
1277 CAATCGTGTTCAGACATGACGTGAAGAGCTGGCTGAATAATACGCCAG 1326
307 HisLeuProValPheGlnHisArgArgProAspLeuTyr 319
|||||: : : : : |||
1327 CAATCCCGCTTTTGTAGACAGAGCGATCAGACCTCTAT 1365
seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI3787
seq_documentation_block:
ID ABLI3787 standard; cDNA; 898 BP.
XX
XX ABLI3787;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 35843.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW
XX
XX pharmaceutical; gene; ss.
OS
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX

```


THIS PAGE BLANK (USPTO)

OM of: US-09-357-675C-21 to: EST.* out_format : pfs
Date: Apr 29, 2002 10:01 AM
About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+g2n.model -DEV=x1h
-O=/usr2.1/USPRO.spool/US09357675/untat_29042002.091756.3146/app.query.fasta.1.1208
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEA_SIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09357675@cgn1.1.6497
-NCPUP=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:
Query: US-09-357-675C-21
Query length: 327
Database: EST*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 2545.650000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Document
gb_hic:AK004988	1479.00	2490.04	1.9e-129	2993	AK004988 Mus musculus adult m
gb_est1:AL520768	1458.00	2468.05	3.2e-128	890	AL520768 LTI_NFL004.NE
gb_hic:AK002269	1458.00	2463.11	6.0e-128	1374	AK002269 Mus musculus adult m
gb_est1:AL529152	1413.00	2391.96	5.5e-124	844	AL529152 LTI_NFL001.NE
gb_est1:AL522844	1357.50	2296.29	1.2e-118	928	B1822844 603040166F1 NIH.MGC.11
gb_est1:AL522844	1308.00	2212.98	5.1e-114	846	B1769604 603054985F1 NIH.MGC.12
gb_est1:AL522844	1298.00	2196.67	4.1e-113	793	B6436616 602888424F1 NIH.MGC.18
gb_est1:AL529163	1293.00	2119.04	8.7e-109	864	AL529163 LTI_NFL008.TC
gb_est1:AL529163	1218.00	2060.75	1.5e-105	761	B6762506 602733954F1 NIH.MGC.49
gb_est1:AL520767	1206.00	2039.66	2.2e-104	992	B6703176 602686082F1 NIH.MGC.95
gb_est1:AL520767	1201.50	2029.62	8.3e-104	992	AL520767 LTI_NFL004.NE
gb_est1:AL522373	1179.50	1992.54	9.7e-102	960	AL522373 LTI_NFL004.NE
gb_est1:AL522373	1169.00	1974.81	9.5e-101	946	B1757823 603030404F1 NIH.MGC.11
gb_est1:AL522373	1135.50	1919.44	1.1e-97	813	B1755493 603027349F1 NIH.MGC.11
gb_est1:AL522373	1115.00	1887.45	6.9e-96	625	B1711300 1d96f1c1.x1 Human insul
gb_est1:AL522373	1109.00	1875.25	3.3e-95	746	B1752623 603028471F1 NIH.MGC.11
gb_est1:AL522373	1097.50	1854.47	4.8e-94	828	B6895409 601434778F1 NIH.MGC.72
gb_est1:AL522373	1089.00	1839.92	3.1e-93	832	B1647584 603278164F1 NIH.MGC.19
gb_est1:AL522373	1087.00	1837.09	4.4e-93	791	B6766506 602663693F1 NIH.MGC.59
gb_est1:AL522373	1079.50	1824.45	4.2e-92	781	B6532265 602613555F1 NIH.MGC.61
gb_est1:AL522373	1077.00	1818.98	4.5e-92	869	B6363013 602828020F1 NCI.CGAP.C
gb_est1:AL522373	1067.50	1803.00	3.5e-91	853	AL522373 602979783F1 NCI.CGAP.C
gb_est1:AL522373	1045.50	1764.89	4.7e-89	900	B6968872 602266558F1 NIH.MGC.84
gb_est1:AL522373	1041.50	1758.55	1.0e-88	863	B1332807 602885469F1 NCI.CGAP.1
gb_est1:AL522373	1035.00	1750.91	2.8e-88	638	B6853807 ux23905.y1 Soarres.thym
gb_est1:AL522373	1025.00	1730.04	4.1e-87	893	B6667049 DRABTC11 Rat DRG librra
gb_est1:AL522373	1013.50	1711.32	4.5e-86	827	B6966074 601852067F1 NIH.MGC.56
gb_est1:AL522373	1006.50	1699.93	2.0e-85	801	B1327600 602979783F1 NCI.CGAP.1
gb_est1:AL522373	1006.00	1702.55	1.4e-85	579	B1711312 1d96f1c1.y1 Human insul
gb_est1:AL522373	1003.00	1693.38	4.5e-85	830	AL580840 603182327F1 NIH.MGC.12
gb_est1:AL522373	1000.50	1691.78	5.5e-85	657	B1916460 601852067F1 NIH.MGC.56
gb_est1:AL522373	999.50	1686.49	1.1e-84	900	B6430308 602495519F1 NIH.MGC.75
gb_est1:AL522373	952.00	1610.92	1.8e-80	561	B1340308 365533 MARC 2P1G Sus s
gb_est1:AL522373	951.50	1605.71	4.4e-80	821	B6780562 601201033F1 NCI.CGAP.4
gb_est1:AL522373	939.00	1586.73	3.9e-79	593	B6392424 601308982F1 NIH.MGC.44
gb_est1:AL522373	931.00	1574.50	1.9e-78	593	B6322101 maa44b02.y1 NCI.CGAP.1
gb_est1:AL522373	912.00	1542.92	1.1e-76	552	B6412111 1f26b04.y1 Melton NORI
gb_est1:AL522373	903.00	1523.53	1.3e-75	789	B6385518 602047063F1 NCI.CGAP.1
gb_est1:AL522373	896.00	1515.97	3.4e-75	537	B6412010 1f26b04.y1 Melton NORI
gb_est1:AL522373	896.00	1509.60	7.7e-75	940	B6135684 601780826F1 NCI.CGAP.1

gb_est1:AL547581 + 894.00 1507.01 1.1e-74 875 | AL547581 LTI_NFL006
gb_est2:BE243444 + 889.50 1499.48 2.8e-74 864 | BE243444 601878153F1 NIH.MGC
gb_est2:BE287010 + 887.50 1499.24 2.9e-74 654 | BE287010 601097244F1 NCI.CGAP
gb_est1:AM956706 + 879.00 1486.20 1.6e-73 576 | AM956706 EST368776 MAGE rese
gb_est2:BI341863 + 862.00 1458.24 5.6e-72 527 | BI341863 369447 MARC 2P1G Su

seq_name: gb_hic:AK004988

seq_documentation block:

LOCUS AK004988 2993 bp mRNA linear HIC 19-JAN-2002

DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched

ACCESSION library, clone:1300013F05:nitrlase 1, full insert sequence.

KEYWORDS AK004988.1 GI:12836590

SOURCE HIC, CAR trapper.

ORGANISM Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
clone:1300013F05.REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (sites)REFERENCE Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253

REFERENCE MEDLINE 10349636

REFERENCE PUBMED 10349636

REFERENCE 2 (sites)

TITLE Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374

REFERENCE MEDLINE 11042159

REFERENCE PUBMED 11042159

REFERENCE 3 (sites)

TITLE Shihata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Tashiro,H., Itoh,M.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashikawa,K.,
Fujiwaka,Y., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913

REFERENCE MEDLINE 11076861

REFERENCE PUBMED 11076861

REFERENCE 4 (sites)

TITLE The RIKEN genome exploration research group phase II team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 2993)

REFERENCE MEDLINE 11076861

REFERENCE PUBMED 11076861

REFERENCE 4 (sites)

TITLE Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Furumoto,Y., Furumoto,M., Hanaoka,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiroaka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasikawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schimml,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Teijima,Y., Toyota,T., Yamamura,T., Yamashita,T.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
Direct Submision
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-res@gsic.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGAGAGAGAGCGCCGCAACTGACATTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adaptor of sequence [5'-
GAGAGAGAGAGAGATCCAGATTCATTTATTTAAACCCCTCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

FEATURES

source

1..2993
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="MGD:MG1:1895066"
/db_xref="taxon:10090"
/clone="1300013F05"
/sex="male"
/tissue_type="liver"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
169..1041
/gene="Nrl1"
169..1041
/gene="Nrl1"
/note="data source:MGD, source key:MG1:1350916,
evidence:ISS
nrl1:ase 1
putative"
/codon_start=1
/protein_id="BAB23723.1"
/db_xref="GI:12836591"
/translation="MSSSTSWELPLVAVCOYTPPNKQENKTCALTYOEARIGACI
APLRAPDFIARNPARETLLSEPLNGDLGQYSQARCGTMSLGGHREGQDEON
QKINCHVILNSKGSVAVSYKRTKLDVEIPGQPMRSNTKPGELTPEPKTPAG
VGLAICYDMREPPELSIKLAQGAELITLYSAFSGTGPAMHEVLLRAVAISQCVIA
AAQGRHNETRASYSKSMVVDPWGTIVARCEGPGCLARIDLHFLQOMRHLPEVQH
RRPDLGSLGHPIS"
2973..2978
/note="putative"
2993
/note="putative"

gene

CDS

polyA_signal
2993
/note="putative"
polyA_site
BASE COUNT 785 a 739 c 747 g 722 t
ORIGIN

alignment_scores:
Quality: 1479.00 Length: 326
Ratio: 4.833 Gaps: 2
Percent Similarity: 93.865 Percent Identity: 84.356

alignment_block:

US-09-357-675c-21 x AK004988 ..

Align seg 1/1 to: AK004988 from: 1 to: 2993

2 LeuGlyPheIleThrArpProHisArgPheLeuSerIleuLeuArgPro 18
|||||
73 CTGGGCTCTACACAGGCTCTCTACACAA.....CTCTGTGTAC 113
18 ogIyleuArgIleProGlnIleuSerValIleuCysAlaGlnProArgPro 35
||| |||
114 CGGATACCGGATTTACTTCGACACCCAGACTTGTACTACACACGAGCCCA 163
35 rgaIaMetAlaIleSerSerSerCysGlnIleuProIleuValAlaVal 51
|||
|||

164 GAACATGATGCC...TCATCACTTCCGGGAGCTGCCCTGTGTGCTGTG 210
52 CysGlnValIleThrSerThrProAspIleGlnGlnAsnPheIleThrCysAl 68
|||||
211 TGCCAGGTAACTACACACCAACCAAGCAAGAGACTTTTAAACATGTGC 260
68 agIyleuValArgGluAlaAlaArgIleuGlnIleuAlaPheLeu 85
|||||
261 TGAGTTGGTTCAAGAGGCTCCAGACTGGGCTGCTGCTGCTGCTGCTGC 310
85 roGluAlaPheAspPheIleAlaArgAspProAlaGlnIleuHisIleu 101
|||||
311 CTGAGCATTTGACTTATTTGACGAACCTTCCGAGACATTACTCTC 360
102 SerGlnProIleuGlnIleuSerIleuGlnIleuGlnIleuValIleu 118
|||||
361 TCCGAACCACTGAATGGGATCTTTTGGGCAATATAGCAGCTTCCAG 410
118 gGluCysGlnIleuThrPheLeuSerIleuGlnIleuHisGlnIleuArgGln 135
|||||
411 GGAATGTGGAATCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 460
135 spITrPGLuGlnIleuThrGlnIleuHisIleuThrAsnIleuValIleu 151
|||||
461 ACTGGGAGGAGAAATCAGAAATCTCAATTTGATGCTGCTTGAACAGC 510
152 LysGlnAlaValAlaValAlaThrTyrArgGlnIleuHisIleuCysAspAla 168
|||||
511 AAGGATCACTAGTGGCCACTTACAGGAACACATCTGTGCATGTAGA 560
168 uIleProGlnIleuGlnIleuProMetCysGlnIleuSerIleuMetProGln 185
|||||
561 GATGCCAGGTGACAGGCGCGATGAGAGAAACACATTTACCAACCTGGAG 610
185 roSerIleuGlnIleuProValSerIleuProAlaGlnIleuHisIleuAla 201
|||||
611 GCACCTTTGAGCCCTGTCACAGACCGGCTGAGGAGTGGTCTGAGCA 660
202 ValCysTyrAspMetIleuPheProGlnIleuSerIleuAlaIleuAlaGln 218
|||||
661 ATCTGTTATGACATGCGGCTTCCCTGAACCTTTTGAATGTGCTCAAGC 710
218 agIyleuAlaIleuThrIleuProSerAlaPheGlnIleuSerIleuThrGln 235
|||||
711 TGGGCGAGAAATACTACTTATCTTACGCTTGGATCTGTACAGGTC 760
235 roIleuHisIleuProGlnIleuValIleuAlaArgAlaIleuGlnIleuHis 251
|||||
761 CAGCCCACTGGAGGAGTGTCTGCGGCGCGCGCATTTGAATCTCAAGTC 810
252 TyrValAlaAlaAlaIleuGlnIleuGlnIleuHisIleuAlaIleuArg 268
|||||
811 TATGTATATGACACACGCGATGTGACGCGCATGCAACCAAGACAGAG 860
268 rTyrGlnIleuHisSerMetValAlaAspProIleuGlnIleuValAlaArg 285
|||||
861 TTATGCGCATGACATGCTGCTGACCCGCGGCGCACAGTGTGCGCGCT 910
285 ysSerGlnIleuProGlnIleuCysIleuAlaArgIleuAspIleuHisIleu 301
|||||
911 GCTCCGAGGAGACAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 960
302 ArgGlnIleuArgIleuHisIleuProValPheGlnIleuHisIleuArgProAsp 318
|||||
961 CAACACATGCGCCACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1010
318 uTyrGlnIleuHisIleuGlnIleuHisIleuSer 327
|||||
1011 CTATGCGAGTCTGCGTCACTGCTGCT 1038
seq_name: gp_est1:AL520768
seq_documentation_block:

```

LOCUS       AL520768                890 bp    mRNA    linear    EST_13-FEB-2000
DEFINITION  AL520768 LTL_NFL004_NBC2 Homo sapiens cDNA clone CSDB002YD05 5
ACCESSION   AL520768
VERSION     AL520768.1  GI:12784261
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
AUTHORS     Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
    Source
        1..890
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CSDB002YD05"
            /clone_lib="LTL_NFL004_NBC2"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DHL0B"
            /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-Oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies, a division of Invitrogen 9800 Medical
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8376 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT   191 a      251 c      241 g      206 t      1 others
ORIGIN

alignment_scores:
    Quality: 1458.00      Length: 277
    Ratio: 5.302          Gaps: 0
    Percent Similarity: 99.278      Percent Identity: 99.278

alignment block:
US-09-357-675C-21 x AL520768 ..

Align seg 1/1 to: AL520768 from: 1 to: 890

1 MetLeuGlyPheIleThraArgProPHisArgPheLeuSerIleuLeu 17
|||||
60 ATGCTGGGGGTGATCATCAGCGCTCTCTCAAGATTCTCTGCTTGtG 109
|||||
17 sPrcGlyLeuArgIlePrcGlnLeuSerValLeuCySalaGlnProArgp 34
|||||
110 TCCTGGAGCTCCGATPACTCAACTCTCAGTACTTGTGCTCAGCCAGG 159
|||||
34 rGArgAlaMetAlaIleSerSerSerSerCysGlnLeuProLeuValAla 50
|||||
160 CCAGAGCCATGGGTATCTCTCTCTCTCTCTGCAATGCCCTGGGtGCT 209
|||||
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelysthrcy 67
|||||
210 GTGTGCGCAGTAATCATCGACGCCAGCAAGCAACAATACTTTAAACATG 259
|||||
67 salagIleuLeuValArgGlnAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
260 TGGTGTAGCTGTTCGAGAGGCTGCCAGACTGGGTGCTGCTGCTTTC 309
|||||
84 euProctIuAlaPheAspPheIleAlaIArgAspProAlaGlnurThrLeuHs 100
|||||

```

310 TGGCTGAGGCAATTGACTTCATTTGGACAGGGAGCCCTCAGAGACGCTACAC 359

101 LeuSerGluProLeuGlyGlyLeuLeuGluGluTyrThrGlnLeuAl 117
|||||
360 CTGTCTGAAACCACTGGGTGGAAACTTTTGGAAAGATTACACCAGCTTGC 409

117 aargGluCysGlyLeuTyrPheLeuSerLeuGlyGlyPheHisGluArgGlyG 134
|||||
410 CAGGGAAATGTGGACTGTGGCTGTCCCTTGGGTGGTTTCCATGAGCCGTGGCC 459

134 LAspPrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
|||||
460 AAGACTGGAGACGACTCAGAAAATTTACAATTTGTACGTGCTGTGAAC 509

151 SerTyrGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspVa 167
|||||
510 AGCAAAAGGGGAGTACTAGTGGCACTTACAGAGACACATCTGTGTACGT 559

167 IgGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
|||||
560 AGAGATTCCAGGGCAGGGGCTATGTGTGAAGCAACTTACATCTCTG 609

184 LProSerLeuGluGlnSerProValSerThrProAlaGlyLysIleGlyLys 200
|||||
610 GGCCACAGTCTTAGTACCTGTCAGCACACACAGCAGGATGGTGTCTA 659

201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
660 GCGTGTGCTATGACATGGGGTTCCTGAAACTCTCTGTGGCATTTGGCTCA 709

217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
|||||
710 AGCTGAGAGAGAGATTACTTACCATTCCTTCAGCTTTGGATCCATTACAG 759

234 LProAlaHisThrPgluValLeuLeuArgAlaArgAlaIleGluThrPgl 250
|||||
760 GCCACCCCACTGGGAGGTGTGCTCGGGCCCTGTCTTTCGAAACCCAG 809

251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluIuysArgAl 267
810 TGGTATGTAGTGGCAGCAGCAGCAGTGTGGACGCCACCATGAGAGAGAC 859

267 aSerTyrGlyHisSerMetValValAspPro 277
|||||
860 AAGTATGGCCACAGCATGTGTGGKAGACCCC 890

seq_name: gb_htc:AK002269

seq_documentation_block: 1374 bp mRNA linear HTC 19-JAN-2002

LOCUS AK002269

DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610007B07:nitrilase 1, full insert sequence.

ACCESSION AK002269

VERSION AK002269.1 GI:12832130

KEYWORDS HTC: CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library

ORGANISM Mus musculus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

MEDLINE Meth. Enzymol. 303, 19-44 (1999)

PUBMED 99279253

10349636

2 (sites)

REFERENCE Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

AUTHORS Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

235 ProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251
|||||
807 CAGGCCACATGGAGGTGCTGGGGGGCCGCCCATTAATCTCGTGTG 856
|||||
251 sTrValValAlaAlaIleGlnCysGlyArgHisHisGlnLysArgAlaAs 268
|||||
857 CTATGTAAATAGCAGACGCCAGTGTGACCCACATGAACAAGACAA 906
|||||
268 eTrArgLysHisSerMetValValAspProTrpGlyThrValAlaIleArg 284
|||||
907 GTTATGGCCATAGCATGTGTGTACCCGTGGGACAGTGTGGCCGC 956
|||||
285 CysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLysAsnTyrLe 301
|||||
957 TGTCTCCAGGAGACCAGGCTCTGCTGCTCGAATTGATCTCCACTTCT 1006
|||||
301 uArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspL 318
|||||
1007 ACAACAGATGCGCCACACCTGCTGTTCACAGACAGACAGCTGACC 1056
|||||
318 eUTrpGlyAsnLeuGlyHisProLeuSer 327
|||||
1057 TCTATGGCAGTCTGGGTCAATCTCTCT 1085
|||||

seq_name: gb_est1.AL529152

seq_documentation_block:

LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5
prime, mRNA sequence.

ACCESSION AL529152
VERSION AL529152.1 GI:12792645
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 844)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

CONTACT: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers

source

1..844

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DD004YM06"

/clone_lib="LTI_NFL001_NBC4"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@life.techn.com URL :
http://fulllength.invitrogen.com"

BASE COUNT

182 a 238 c 235 g 189 t

ORIGIN

alignment_scores:

Quality: 1413.00

Ratio: 5.373

Percent Similarity: 100.000

Length: 263

Gaps: 0

Percent Identity: 100.000

alignment_block:
US-09-357-675c-21 x AL529152 ..
Align seq 1/1 to: AL529152 from: 1 to: 844

65 LysThrCysAlaGluLeuValArgGluAlaIleArgLeuGlyAlaCysLe 81
|||||
1 AAAACATGCTGCTGAGCTGTGTTGAGAGGCTGCCAGACTGGGCTGCTGCT 50
|||||
81 uAlaPheLeuProGluAlaPheAspPheIleAlaIleArgAspProAlaGlu 98
|||||
51 GGCCTTCTCTGCTGAGGCAATTCATTCATTCAGCAGGAGCCCTCAGAGA 100
|||||
98 hTrLeuHisLeuSerGluProLeuGlyLysLysLeuGluGluGluTyrThr 114
|||||
101 CGCTACACCTGTCTGAACCACTGGGTGGGAAATTTGGAAATACACACC 150
|||||
115 GlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGln 131
|||||
151 CAGCTTCCAGGGAATGTGGACTGTGCTGTGCTGTGGGTGTTCCATGA 200
|||||
131 uArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValL 148
|||||
201 GCGTGGCCAAAGACTGGGAGCAGACTCAGAAATCTACAAATGTGCACGTGC 250
|||||
148 eUeUAsnSerLysGlyAlaValValAlaIleThrTyrArgLysThrHisLeu 164
|||||
251 TGTGGAACAGCAAAAGGGGAGTAGTGGCCACTTACAGGAACACATCTG 300
|||||
165 CysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerTh 181
|||||
301 TGTACCTAGAGATTCCAGGCGAGGCGCTATGTGTAAACCACTCTAC 350
|||||
181 mEtProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 198
|||||
351 CATGCTGGGCGCCAGCTGTGAGTCACTGACCTGCAGCAGACAGCAGGCAAGA 400
|||||
198 leGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAla 214
|||||
401 TTGCTTACGCTGTCTGTATACATGCGGTCCCTGAACCTCTCTGCGCA 450
|||||
215 LeuAlaGlnAlaGluAlaGluIleLeuThrTyrProSerAlaPheGlySe 231
|||||
451 TTGGCTCAAGCTGAGCAGAGATCTTACCTATCTTCAGCTTTTGATC 500
|||||
231 rIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaIleG 248
|||||
501 CATACAGGCCACAGCCACTGGAGGTGTGCTGCGGCGCCGTGCTATGC 550
|||||
248 lUTrpGlnCysTyrValValAlaAlaIleGlnCysGlyArgHisHisGlu 264
|||||
551 AAACCCAGTGTATGTAGTGGCAGCAGCAGCATGTGAGCGCCACCATGAG 600
|||||
265 LysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVa 281
|||||
601 AAGAGACCAAGTTATGGCCACAGCATGTGTATACCCCTGGGAAACAGT 650
|||||
281 lValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgTLeAspL 298
|||||
651 GGTGGCCCGCGTCTGTGAGGAGGCCAGGCTGTGCTGCCCGGAAATGACC 700
|||||
298 eUAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArg 314
|||||
701 TCAACTATCTGCGACAGTTCGGCCGAGACCTGCTGTGTCCACACCCGC 750
|||||

seq_name: gb_est2.BI822844

seq_documentation_block:

LOCUS BI822844

928 bp

mRNA

linear

EST 04-OCT-2001

DEFINITION 603040166F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5180897 5', mRNA sequence.
ACCESSION BI822844
VERSION BI822844.1 GI:15934394
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 928)
NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L14M11451 row: d column: 18
High quality sequence stop: 822.
Location/Qualifiers
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5180897"
/clone_1db="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector: PCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."
BASE COUNT 198 a 256 c 259 g 214 t 1 others
ORIGIN
alignment_scores:
Quality: 1357.50 Length: 295
Ratio: 4.848 Gaps: 4
Percent Similarity: 94.915 Percent Identity: 93.559
alignment_block:
US-09-357-675C-21 x BI822844 ..
Align seg 1/1 to: BI822844 from: 1 to: 928
1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeu 17
|||||
39 ATGCTGGGCTTCATCACAGGCTCCTCCACAGATTCCTGTCCTTGTG 88
|||||
17 SPtGGIleuArgIleProGlnIleuSerValLeuGlySalGlnProArgp 34
|||||
89 TCGTGAAGTCGGATACCTCACTCAGACTTGTGCTCAGCCAGCC 138
|||||
34 roArGAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50
|||||
133 CAGAGCCAGCATGCTATCTCTCTCTCTCCGCACTGCCCTGCGTGT 188
|||||
51 ValCysGlnValInhrSerThrProAspLysGlnGlnAsnPhelysthrcy 67
|||||
189 GGTGTCCAGGTACATGACGACGACAGACAGCAAGACTTTAAACATG 238
|||||
67 sAlaGluLeuValArgGluAlaIleArgLeuGlyAlaCysLeuAlaPheL 84
|||||

239 TCGTAGAGCTGGTTCGAGAGGCTGCCAGACTGGCTGGCTGGCTGGCTTTC 288
84 eupProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
289 TGCCTGAGGCAATTTGACTTCATTGSCAGGAGCCCTGCAGAGAGCTACAC 338
|||||
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluGluThrGlnLeuAl 117
|||||
339 CTGTCTGAACCACTGCGGTGGGAACTTTTGAGAGATACACCCAGAGCTTC 388
|||||
117 aArgGluCysGlyLeuThrPheLeuSerLeuGlyGlyPheHisGlnArgLys 134
|||||
389 CAGGAGATGTGAGCTGTGGCTTGTGGGTGTTCATGAGCGGTGGC 438
|||||
134 InAspTPrGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
|||||
439 AAGAGCTGGAGCAGACTCACAANAATGTACAAATGTCTACGCTGCTGAC 488
|||||
151 SerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeuGlyAspVa 167
|||||
489 AGCAAGGGGCGAGTACGTGGCCACTTACAGAGAACACATCTGTGTACGT 538
|||||
167 IGluIleProGlyGlnGlyPrometCysGluSerAsnSerThrmProG 184
|||||
539 AGAGATTCACAGCAG. GGGCTATGTGTGAAGCACTTACCATGCTGT 587
|||||
184 IyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
588 GGGCCAGCTCTTGAAGTCACTGTGACGCCACCCAGCAGGCAAGATTGCTTA 637
|||||
201 AlValAlcysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
|||||
638 GCTGTGCTATGATGACATGCGGTCCTCCGAGACTCTCTGCAATGGCTCA 687
|||||
217 naIaGlyAlaGluIleLeuThr TyrProSerIaPheGlySer... Ile 232
|||||
688 AGCTGAGGAGAGATCTTACCTTATCTTACCTTGTGGATCCCATATA 737
|||||
233 ThGlyProAlaHisTPrGluValLeuLeuAlaArgAlaIleGluThr 249
|||||
738 CAGGGCCACCCACACGTGGAGGTGTGCTCGGGGNCCTGCTATTCGAAC 787
|||||
249 rGlnCysTyrVal. ValAlaAlaAlaGlnCysGly. ArgHisHisGlu 265
|||||
788 CCAGTCTGATGTATGATGGCCGACACAGCTGTGAACGCCACCATGAGAA 837
|||||
265 sArgAlaSerTyrGlyHisSerMetValValAspProTPrGlyThrVal 282
|||||
838 GAGAGC. AGTTATGGGCACAGCATGGGTGTAACCCCTGGGGGACAGGGG 886
|||||
282 alAlaArgCysSerGluGlyProGlyLeu 291
|||||
887 CCCGTGCTGTGAGGCGCAGGCTGTGGCTTG 915
|||||
seq_name: gp_est2:BI769604
seq_documentation_block:
LOCUS BI769604 846 bp mRNA linear EST 25-SEP-2001
DEFINITION 603054985F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5', mRNA sequence.
ACCESSION BI769604
VERSION BI769604.1 GI:15761182
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 846)
NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1AM1512 row: 3 column: 22
 High quality sequence stop: 838.

FEATURES

source

1..846

Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5204469"
 /clone_lib="NIH_MGC_122"
 /lab_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH MGC Library."

BASE COUNT 179 a 240 c 226 g 201 t
 ORIGIN

alignment_scores:

Quality: 1308.00 Length: 264
 Ratio: 4.992 Gaps: 2
 Percent Similarity: 99.242 Percent Identity: 99.242

alignment_block:

US-09-357-675c-21 x B1769604 ..

Align seg 1/1 to: B1769604 from: 1 to: 846

1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeu 17
 |||||||
 59 ATGCTGGGCTTCATACACAGGCTCTCCACAGATCTGCTTCGTG 108
 |||||||
 17 SPGGLYLeuArgIleProGlnLeuSerValLeuGlyAlaGlnProArg 34
 |||||||
 109 TCCTGGACTCCGGATACCTCACTCAGACTTGTGCTCACGCCAGG 158
 |||||||
 34 roArGAlaMetAlaIleSerSerSerSerSerGluLeuProLeuValAla 50
 |||||||
 159 CCAGAGCCATGGCTATCTCTCTCTCTCCGCACTGCCCTGGTGGCT 208
 |||||||
 51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelThrC 67
 |||||||
 209 GTGTGCAGTACATCGACGACGACCAACAGACATTAAACATG 258
 |||||||
 67 salagLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
 |||||||
 259 TGCTGGAGTGGTTCGAGAGGCTCCAGACTGGGTGCTGCTGCTTCC 308
 |||||||
 84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
 |||||||
 309 TGCCTAGGATTTGACTTATTCACAGGAGCCCTGCAGAGAGCTACAC 358
 |||||||
 101 LeuSerGluProLeuGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
 |||||||
 359 CTGTCAACACCACTGGGTGGAAACTTTGGAAAGATACACCACTTGC 408
 |||||||
 117 aaArgLysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGly 134
 |||||||
 409 CAGGAGATGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 458
 |||||||
 134 InaSPTrpGluGlnThrGlnLysIleTyrAsnGlyHisValLeuLeuAsn 150

seq_name: gb_est2:BG436916

seq_documentation_block:

LOCUS BG436916 793 bp mRNA linear EST 14-MAR-2001
 DEFINITION 602488424F1 NIH_MGC_18 Homo sapiens CDNA clone IMAGE:4620892 5',
 mRNA sequence.
 ACCESSION BG436916
 VERSION BG436916.1 GI:13343422
 KEYWORDS EST.

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 793)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DTF/Gazdar
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1CMI382 row: 0 column: 05
 High quality sequence stop: 767.

FEATURES

source

1..793

Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4620892"
 /clone_lib="NIH_MGC_18"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and


```

|||||
178 GTGTGCAGTACATCGACGCCAGAAVACAATTAACATG 227
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
228 TGTGTAGCTGTTGAGAGAGCTGCCACACGTGGGCTGCTGCTTCC 277
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
278 TGCTGTAGGACATTTGACTTATTCATGGGACCCCTGCAGAGACCTTACAY 327
101 LeuSerGluPro..LeuGlyGlyLeuLeuGluGluTyrThrGlnLeuA 117
|||||
328 CTGTCTGAACCTACTGGGTGGGAAACTTTTGAAGATACATACACAGCTK 377
117 laArgGlu..CysGlyLeuTyrPleuSerLeuGlyGlyPheHisGluArgL 133
|||||
378 CCAAGGATATGTGACTCTKCTGCTTCTGTGGGTGGTTCATGAGCTGG 427
133 yGlnAspTrpGluGlnThrGlnLysIleTyrAsnGlyHisValLeuLeuA 150
|||||
428 CTAAGACATGGGACAGACTYAGAAATCTACAATGTCACTGCTGCTGA 477
150 snSerLysGlyAlaValAlaAlaThrTyrArgLysThrHisLeuGlyAsp 166
|||||
478 ACAGCAAGGGCAGTAGTGGCCACTTACAGGAAGACATCTGTGTGAC 527
167 ValGluLeuProGlyGlnGlyProMetCysGluSerAsnSerThrMetPr 183
|||||
528 GTTGAATATTCAGGCGAGGGCCTATGTGAAAGCAACTTCAACATGCC 577
183 ogLysProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyL 200
|||||
578 TGGGCCCATGCTTGTAGTACACCTGTGCACACACAGGAGGAGATGGTGC 627
200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
|||||
628 TACTGTCTGTATGACATCGGCTGCCCTGCACCTCTCTCRCATTGGCT 677
217 GlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIle..T 233
|||||
678 CMACTGGGCGCCAGAGAACTTMACTTMACTTMACTTMACTTMACTTMA 727
233 hrcGlyProAlaHisTrp..GluValLeuLeuArgAlaArgAlaIleGluTh 249
|||||
728 CAGGCCACCCCACTGGGCGAGGTGTGCTGGGCCCTGCTCWCAGAAAM 777
249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
|||||
778 CCACTGCTTMMGTAGTGCWGCACACAGCTGTGACGCCACMATGACAGA 827
266 rGlnAspTyrGlyHisSerMetValValasp 276
|||||
828 GAGCAAGTTATGGCCACACATGATGTGTAGAC 859
seq_name: gp_est12.BG762506

seq_documentation_block:
LOCUS BG762506 764 bp mRNA linear EST 15-MAY-2001
DEFINITION 603733954F1 NIH_MGC_49 Homo sapiens CDNA clone IMAGE:4859292 5',
RNA sequence.
ACCESSION BG762506
VERSION BG762506.1 GI:14073159
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC).
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

```

Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC/DCPD/DFP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1714 row: 1 column: 13
High quality sequence stop: 749.
FEATURES
source
1..764
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4859292"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 162 a 218 c 201 g 183 t
ORIGIN
alignment_scores:
Quality: 1218.00 Length: 240
Ratio: 5.183 Gaps: 2
Percent Similarity: 97.917 Percent Identity: 97.917
alignment_block:
US-09-357-675C-21 x BG762506 ..
Align seq 1/1 to: BG762506 from: 1 to: 764
1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuGly 17
|||||
49 ATGCTGGGCTTCAACACAGGCTCTCCACAGATTCTGTCCCTTCTGTG 98
|||||
17 sProGlyLeuArgIleProGln..LeuSerValLeuGlyAlaGlnProArg 33
|||||
99 TCTGGACTCCGGATACCTCTAATCTCAGTACTTGTGGTCCAGCCAGG 148
|||||
34 ProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAl 50
|||||
149 CCCAGACCAATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 198
|||||
50 aValCysGlnValThrSerThrProAspLysGlnGlnAspPheLysThrC 67
|||||
199 TGTGTGCCAGGTATACATCGACGCCAGACACACAGAACTTTAAACAT 248
|||||
67 ySalGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 83
|||||
249 GTGTGTAGCTGTTGAGAGAGCTGCCACAGACTGGTGGCTGCTGCTTTC 298
|||||
84 LeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuH 100
|||||
299 CTGCTGAGGACATTTGACTTATTCATGGGAGGCCCTGCAGAGAGCTTACA 348
|||||
100 sLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuA 117
|||||
349 CTGTCTGAACCTACTGGGTGGGAACTTTTGAAGATACATACACAGCTTG 398
|||||
117 laArgGluCysGlyLeuTyrPleuSerLeuGlyGlyPheHisGluArgGly 133
|||||
399 CCAAGGATATGTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 448

```

```

134 GlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAs 150
|||||
449 CAAGACTGGGAGCAGACTCAGAAAATCTACAAATTGCACGTGCTGTGA 498
|||||
150 nserLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAsp 167
|||||
499 CAGCAAAAGGGCGAGTAGTGCCCTTACAGGAACACATCTGTGACG 548
|||||
167 AlGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPro 183
|||||
549 TAGAGATTCCAGCGCAGGGGCTATGTGTGAAGCAACTCTACCATGCC 598
|||||
184 GlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLe 200
|||||
599 GGCCCAAGCTTGTGACCTGTGCACACACGAGGCAAGATTGGTCT 648
|||||
200 uAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 217
|||||
649 ACCTGCTCTATGACATCGGTTCCCTGAACTCTGTGGCATTTGGCTC 698
|||||
217 InAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233
|||||
699 AAGCTGACACAGAGATCTTACCTATCTTCAGCTTTGGATCCATTACA 748
|||||
234 GlyPro. AlaHisTrp 238
|||||
749 GCCCAGGCCCACTGG 764

```

seq_name: gb_est2:BG703176

seq_documentation_block:

LOCUS BG703176 790 bp mRNA linear EST 07-MAY-2001
 DEFINITION 602686082P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4818405 5',
 mRNA sequence.

ACCESSION BG703176
 VERSION BG703176.1 GI:13975252

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 790)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-tremail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNLN at:

http://image.lnl.gov

Plate: LRAM10720 row: d column: 22

High quality sequence stop: 766.

FEATURES

SOURCE

```

1..790
Location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:4818405"
  /clone_1lb="NIH_MGC_95"
  * /tissue_type="hippocampus"
  /lab_host="DH10B"
  /note="Organ: Brain; Vector: pBluescriptR (modified
  pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gcgag
  ); Oligo-dt primed using primer 5'-TTTTTATTTTATTTT-3',
  size-selected for average insert size 2.5 kb and
  normalized to ROP 5. This is a primary library enriched
  for full-length clones and constructed using the
  Cap-trapper method (Carninci, in preparation). Library
  constructed by M. Brownstein (NIH/NHGRI, National

```

BASE COUNT 167 a 228 c 206 g 189 t
 ORIGIN Institutes of Health). Note: this is a NIH-MGC Library."

alignment_scores:
 Quality: 1206.00 Length: 238
 Ratio: 5.132 Gaps: 2
 Percent Similarity: 98.739 Percent Identity: 97.899

alignment_block:
 US-09-357-675c-21 x BG703176 ..

Align seg 1/1 to: BG703176 from: 1 to: 790

```

1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeu 17
|||||
79 ATGCTGGGCTTCATTCACACAGGCTCTCTCACAGATTCCTGCTCTGTG 128
|||||
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArg 34
|||||
129 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTGTGCTCAGCCAGG 178
|||||
34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50
|||||
179 CCAGAGCCATGGCTATCTCTCTCTCTCCCTGCGAACCTGCCCTGGCT 228
|||||
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThr 67
|||||
229 GTGTGCAGGTAACATCGACGCGCAGAACACACAGAACATTAAACATG 278
|||||
67 saLaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
|||||
279 TGCTGAGCTGGTTCCGAGAGGCTCCAGACTGGGTGCTGCTGCTTTC 328
|||||
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
329 TGCTGAGGCAATTGACTTCATTGACAGGGGACCTGCAGAGAGCTAAC 378
|||||
101 LeuSerGluProLeuGlyGlyLysLeuGluGluGlyTyrThrGlnLeu 117
|||||
379 CGGTCGAACCACTGGGTGGGAACTTTGGAGAAATACACCCAGCTTCC 428
|||||
117 aArgGlyCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGly 134
|||||
429 CAGGAAATGGAGCTGTGGCTGTGGTGGTGTTCATGAGCATGGCC 478
|||||
134 InAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAs 150
|||||
479 AAGACTGGGAGCAGACTCAGAAAATCTACAAATGTGCACATGCTGGAAC 528
|||||
151 SerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAsp 167
|||||
529 AGCAAAAGGGGAGTAGTGCCACTTACAGGAACACACATCTGTGATGCT 578
|||||
167 IgLuIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPro 184
|||||
579 AGAGATTCCAGGCGCAGGGGCTATGTGTGAAGCAACTCTACATGCTCG 628
|||||
184 LyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLe 200
|||||
629 GGCCAGCTCTTGAAGTCACTGACGACACACAGCAGCAAGATTGGTCTA 678
|||||
201 AlaValCysTyr. AspMetArgPheProGluLeuSerLeuAlaLeuAla 217
|||||
679 GCGTCTGTATTGACATCGGTTCCCTGAACCTCTCTGGCATTTGGCTC 728
|||||
217 InAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIle. Th 233
|||||
729 AAGCTGACACAGATCTTACCTATCTTCAGCTTTGGATCCATTAC 778
|||||
233 relProAla 236
|||||

```

```

779 AGGCCAGACC 788
seq_name: gb_est1:AL520767
seq_documentation_block:
LOCUS      AL520767                992 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION AL520767 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 3
prime, mRNA sequence.
ACCESSION  AL520767
VERSION     AL520767.1  GI:12784260
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 992)
AUTHORS     Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source      1..992
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DB002YD05"
            /clone_1ib="LTI_NFL004_NBC2"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by life technologies. Contact: Feng Liang life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT  229 a      267 c      268 g      222 t      6 others
ORIGIN

```

```

|||||
786 GGAGACACATCTGTGTACGTAGAGATTCCAGGGGAGGGCCATATGCT 737
|||||
177 GUSerAnsSerThrMetProGlyProSerLeuGluSerProValSerTh 193
|||||
736 GAAAGCAACTCTACCATCCCGGGGCCGCTTGTAGTCACCTGTCAGCAC 687
|||||
193 rProAGlyLysTLeuAlaValCysTyrAspMetLrpPheProG 210
|||||
686 ACCAGCAGGCAAGATTGGCTAGCTGTCTGCTATGACATGGCGTCCCTG 637
|||||
210 LuleSerLeuAlaLeuAlaGlnAlaGluLeuThrTyrPro 226
|||||
636 AACCTCTCTGCAFTGGCTCAAGCTGGAGAGATATCTTACCTATGCT 587
|||||
227 SerAlaPheGlySerLleThrGlyProAlaHisTrpGluValLeuLeuAr 243
|||||
586 TCAGCTTTTGATCCATTATACAGGCCACCCAGCTGGAGGTGTGCTCG 537
|||||
243 GAlArGAlaTllegLurThrGlnCysTyrValValAlaAlaGlnCys 259
|||||
536 GGCCCGTGTATCGAACCAGTGTATGTAGTGCAGCAGCAGTGT 487
|||||
260 GYArGHisGluLysArgAlaSerTyrGlyHisSerMetValAlaVal 276
|||||
486 GGAGCCCAACCATGAGAGAGCAAGTATATGCCACACCATATGCTGTA 437
|||||
276 pProTrpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysL 293
|||||
436 CCCCTGGGAGACAGTGTGGCCCTGCTGTAGGGGCCAGCCCTCTGCC 387
|||||
293 euAlaArgLleAspLeuAnsTyrLeuArgGlnLeuArgArgHisLeuPro 309
|||||
386 TTGGCCGATAGACCTCAACTATCTGCAGCAGTTGGCCGACCTGCT 337
|||||
310 ValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHisProLe 326
|||||
336 GTGTTCAGACACCCAGGCTGACCTTATGCAATCTGGGTACCCACT 287
|||||
326 uSer 327
|||||
286 GTCT 283

seq_name: gb_est1:AL522373
seq_documentation_block:
LOCUS      AL522373                960 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION AL522373 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB008YMO2 3
prime, mRNA sequence.
ACCESSION  AL522373
VERSION     AL522373.1  GI:12785866
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 960)
AUTHORS     Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source      1..960
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DB008YMO2"
            /clone_1ib="LTI_NFL004_NBC2"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"

```

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (41) 301 610 8371 Email: liang@lifetech.com URL: <http://fulllength.invitrogen.com>"

BASE COUNT 225 a 259 c 255 g 217 t 4 others
ORIGIN

alignment_scores:
Quality: 1179.50 Length: 231
Ratio: 5.151 Gaps: 2
Percent Similarity: 99.134 Percent Identity: 98.268

alignment_block:
US-09-357-675C-21 x AL522373/rev ..

Align seg 1/1 to reverse of: AL522373 from: 1 to: 960

```

98 ThrLeuHisLeuSerGluProLeuGlyLysLeuLeu.GluGluTyrT 114
|||||
959 ACCGTACACCTGCTCTGAACCACTGGTGGAAC...TTTGAAGAATACA 913
|||||
114 hrgInleuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHis 130
|||||
912 CCCAGCTTGCAGGGAAATGTGGACTGTGGCTGCTGGGGGTTCCAT 863
|||||
131 GluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisVa 147
|||||
862 GAGCGTCCCAAGCTGGAGCACTCAGAAATCTACAAATTTCCAGT 813
|||||
147 lLeuLeuAsnSerLysGlyAlaValAlaThrTyrArgLysThrHisL 164
|||||
812 GCTGCTGAACAGCAAGAGGGCAGTAGTGGCCACTTACAGGAACACATC 763
|||||
164 eucYasAspValGluIleProGlyGlnGlyPrometCysGluSerAsnSer 180
|||||
762 TGTGTGACGTAGAGATTCCAGGCGAGGGCCCTATGTGTGAAGCACTCT 713
|||||
181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyL 197
|||||
712 ACCATGCTGGGCCCACTTTGACTACCTGTGACGACACACGACGACAA 663
|||||
197 sileGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuA 214
|||||
662 GATTGGTCTAGCTGTCTGTATGACATGCG..TTCCTGAACCTCTCTCG 614
|||||
214 lalaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 230
|||||
613 CATTGGCTCAAGCTGGAGACAGATTACTTACTTCCCTTACGCTTTTGA 564
|||||
231 SerIleThrGlyProAlaHisThrGluValLeuLeuArgAlaArgAlaI 247
|||||
563 TCCATTACAGGCCCAAGCCCACTGGAGGTGTGTGGGGCCCTGCTAT 514
|||||
247 eGluThrGlnCysTyrValAlaAlaAlaGlnCysGlyArgHisHisG 264
|||||
513 CGAAACCCAGTGTATGTAGTGGCAGCAGCACTGTGAGCGCCACCATG 464
|||||
264 luyAsArgAlaSerTyrGlyHisSerMetValValaAspProTrpGlyThr 280
|||||
463 AGAAGAAGAGAAATTATGGCCACAGCATGTGTAGACCCCTGGGAGACA 414
|||||
281 ValValAlaArgCysSerGluGlyProGlyLeuGlyLeuAlaArgIleAs 297
|||||
413 GTGTGGCCCGCTGCTGTGAGGGGCCAGGCTGTGCTTGGCCCAATAGA 364
|||||
297 pleuAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisA 314

```

```

|||||
363 CCTCACTATCTGGACAGATTGGCCGACACACTGCTGTGTCCAGACCC 314
|||||
314 rGATrGProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
313 GCAGGCTTGACCTTATGGCAATCTGGGTGATCCACCTGTCT 273

```

seq_name: gb-est2:BI757823

seq_documentation_block:

LOCUS BI757823 946 bp mRNA linear EST 25-SEP-2001
DEFINITION 603030404P1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200560 5',

ACCESSION BI757823

VERSION BI757823.1 GI:15749401

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 946)
NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

plate: LAM11502 row: h column: 01

High quality sequence stop: 708.

Location/Qualifiers

1..946

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5200560"

/clone_lib="NIH_MGC_114"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: EcoRV (destroyed); RNA source anonymous pool of 6

male brains, age range 23-27 yo. Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.5 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 019. Note:

this is a NIH_MGC Library."

alignment_scores:
Quality: 1169.00 Length: 291
Ratio: 4.395 Gaps: 7
Percent Similarity: 91.409 Percent Identity: 83.849

alignment_block:
US-09-357-675C-21 x BI757823 ..

Align seg 1/1 to: BI757823 from: 1 to: 946

```

1 MetLeuGlyPheIleThrArgProTrpHisArgPheLeuSerLeuLeu 17
|||||
69 ATGCTGGCTTCAATCACAGGCCCTCCACAGATTCTGTCTCTCTGTG 118
|||||
17 sProGlyLeuArgGlnProGluLeuSerValLeuCysAlaGlnProArgP 34
|||||
119 TCTGAGCTCCGGATTAACCTCACTCAGTACTTGTGTCAGGCCAGGC 168

```



```
125 GCGGGCCCGTGTATCGAAACCCAGTGTCTATGTAGTGGCACAGCACAGT 76
259 ysglyArgHishglulysArgAlaSerTyrglyHisSerMetValVal 275
75 GTGAGCGCCACCATGAGAGAGAGCAAGTATAGGCCACAGCATGTGTGTA 26
276 AspProTrpGlyThrValValAla 283
25 GACCCCTGGGGACACAGTGTGTGCC 2
```

THIS PAGE BLANK (USPTO)

OM of: US-09-357-675C-21 to: Issued_Patents_NA:* out_format : pfs
Date: Apr 29, 2002 10:02 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+p2n.model -DEV-x1h
-Q=/cgn2_1/USPTO.spool/US09357675/Runat_29042002_091757_3210/app.query.fasta_1.1208
-DB-Issued_Patents_NA -OPMT-fastlap -SUFFIX=p2n.rn1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -RGAPOP=6.000 -RGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=biosubm2 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09357675_@CGN1.118 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-09-357-675C-21

Query length: 327

Database: Issued_Patents_NA:*

Database sequences: 383533

Database length: 122816752

Search time (sec): 85.140000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-63 +			206.00	364.19	1.3e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-63 +			206.00	364.19	1.3e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-25 +			205.00	362.25	1.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-25 +			205.00	362.25	1.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-5 +			204.00	360.31	2.2e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-5 +			204.00	360.31	2.2e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-5 +			204.00	360.31	2.2e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-5 +			204.00	360.31	2.2e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-61 +			203.00	358.37	2.8e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-61 +			203.00	358.37	2.8e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-9 +			202.00	356.43	3.6e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-9 +			202.00	356.43	3.6e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-11 +			202.00	356.43	3.6e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-11 +			202.00	356.43	3.6e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-07-917-111-1 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-479-638-1 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-1 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-1 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-3 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-3 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-7 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-7 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-13 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-13 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-15 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-15 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-17 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-17 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-29 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-29 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-31 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-31 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-47 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-47 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-49 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-49 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-51 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-51 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-65 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-294-871A-65 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-1 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-1 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-3 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-3 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-7 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-7 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-13 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-13 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-15 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-15 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-876-398A-17 +			201.00	354.49	4.7e-12 1785
/cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-876-398A-17 +			201.00	354.49	4.7e-12 1785

seq_name: /cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-294-871A-63

seq_documentation_block:

Sequence 63, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanda, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 340078/1992
 FILING DATE: 21-DEC-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Wegner, Harold C.
 REGISTRATION NUMBER: 25,258
 REFERENCE NUMBER: 74129/127/AOPA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 63:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1785 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 ORIGINAL SOURCE:
 STRAIN: JM109 PAD469
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 233..1144
 US-08-294-871A-63

alignment_scores:
 Quality: 206.00 Length: 340
 Ratio: 1.280 Gaps: 15
 Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-63 ..

Align seg 1/1 to: US-08-294-871A-63 from: 1 to: 1785

```

50 AlavAlcysGlnValThr.....SerThrProAs 59
   |||:||||| :|||:|
164 GCGATGTGTTTCACAAGCTTTCCCGCGCGCTGGCCGACATCCACTAG 213
   |||:||||| :|||:|
59 pLysGlnGlnAspNheLysThr..... 66
   |||:||||| :|||:|
214 GAAGGAGCAGAGGTTTCATGACACGTGATGATCTTGCAGTGGGACAC 263
   67 .....Cys 67
264 AAGGTCCGATCCGCGCGGAGACACGACGACGATGCTGCTGCTT 313
   |||:||||| :|||:|
68 AlagLleuValArgLualAalArgLeuGlyAlaCysLeuAlaPheLe 84
   |||:||||| :|||:|
314 CTCGACATGCTGACGAAGCCGCGACCGCGGCGAATTTCATTGCTT 363
   |||:||||| :|||:|
84 upTgGlu.....AlaPheAspPheIleAlaArg.....Asp 95
   |||:||||| :|||:|
364 CCCGGAACTCGCGCTTACGACCTTCTCCCGCGCTGGCATTTACCGAC 413
   |||:||||| :|||:|
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
   |||:||||| :|||:|
414 AGCGCGAGCTCGATAGCTTCTATGACACCGAATGCCCGCGCGTGGTC 463
   |||:||||| :|||:|
111 GluGluTyrThrGlnLeuAlaArgLysGlyLeuThrLeuSerLeuG 127
   |||:||||| :|||:|
464 CGTCCACTCTTTGAGAAGCCGCGGACACGCGGATCGGCTTCAATCGG 513
   |||:||||| :|||:|
127 yGluPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141
   |||:||||| :|||:|
514 CTACGCTGAACTCGTGTCGAAGCGGC.....GTCAAGC 548
   |||:||||| :|||:|
141 ySileTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
   |||:||||| :|||:|
549 GTCCGCTTCAACAGCGCTATTGCTGATTAAGTACAGCAAGATCGTCGG 598
   |||:||||| :|||:|
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 174
   |||:||||| :|||:|

```

```

599 AAGTATCGTAAGATTCATTTC.....CCGGGTCAACAG.. 631
174 omecYsGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
   |||:||||| :|||:|
632 .....GAGTACGAGGCGCTACCGCGCGTTCACGATCTTGAAAGCGT 674
   |||:||||| :|||:|
191 alSerThrPro.....AlaGly 196
   |||:||||| :|||:|
675 ATTTGAGCCGCGGATCTCGGCTCCGGCTATGACGTGACGCCGCG 724
   |||:||||| :|||:|
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213
   |||:||||| :|||:|
725 AAATGGGAGATTTCATCTGCAACGATCGCGCTGGCTGAGCTGGCG 774
   |||:||||| :|||:|
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 230
   |||:||||| :|||:|
775 GGTATGGCGCTCAGGGCGCGCGAGATCATCTGC.....G 809
   |||:||||| :|||:|
230 LysSerIleThrGlyProAlaHisTrpGluValLeu..... 241
   |||:||||| :|||:|
810 GCGGCTACACACGCGCGACACATCCATGTGTCCCGCAGCAGCACAC 859
   |||:||||| :|||:|
242 .....LeuArgAlaArgAlaIleGluThr 249
   |||:||||| :|||:|
860 CTGACGCTCTCCACCATCTCTATGATGACAGCGCGGTCTTATCGAA 909
   |||:||||| :|||:|
249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisGluLysA 266
   |||:||||| :|||:|
910 CCGGGCGCTGTCGCGCGCGCGGCAAGTGGCGATG...GAGGAGACT 956
   |||:||||| :|||:|
266 rGlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVal 282
   |||:||||| :|||:|
957 GCATGCTGCTCGGCGCACCTCTGATCTGCGCGCGCGGGAAGTCTC 1006
   |||:||||| :|||:|
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
   |||:||||| :|||:|
1007 GCTCTCACTACGACCGCTGGAAGACGAGGTGATCACCGCCCGCTCA 1056
   |||:||||| :|||:|
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
   |||:||||| :|||:|
1057 CGATCGCTCGCGGAGACTGCGTGAACACATCTTCACTTCAACAGCAT 1106
   |||:||||| :|||:|
314 rGArgProAspLeuTyrGly 320
   |||:||||| :|||:|
1107 GTCAGCCCCAGCATTAAGT 1126
   |||:||||| :|||:|
seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-876-398A-63
seq_documentation_block:
; Sequence 63, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiko
; APPLICANT: NAKABA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

CURRENT APPLICATION DATA
APPLICATION NUMBER: US/08/876.398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/0110D
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5500
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 pmd469
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-63

```

alignment_scores:		
Quality:	206.00	Length: 340
Ratio:	1.280	Gaps: 15
Percent Similarity:	47.353	Percent Identity: 25.294

alignment_block:

US-09-357-675C-21 X US-08-876-398A-63

Align seg 1/1 to: US-08-876-398A-63 from: 1 to: 1785

[illegible]

```

111 GUGUGUUTYThrcInLeuAlaIrgLcYgLYleuTrpleuSerLeuG 127
    :::      :::      |||      |||      |||      |||
464 CGTCCACTCTTTAGAAAGCGCGGAACTCGGGATCGCTCAATCTGG 513
127 YGLYPhEhIS.....GLuArGsgLYGlnAsPTrpGLnIothrcInL 141
    |      :::::      |||      |||      :::::      |
514 CTACGCTGAACCTGCTCGAAGCGGC.....GTCAAGC 548
141 ysIleTyraSncYSnHSIvalLeuInsSerYsgLYAlaIvalAla 157
    :::::      |||      |||      |||      |||      |||
549 GTGCGTTCAACAGCTGCATTTTGGTGATTAAGTCAGGCAAGATCGTGGC 598
158 ThTYrArGLysThrHSIleuCYsAsPValGluLeuProGLyGlnGLYpr 174
    |||||      |||||      |||||      |||||      |||||
599 AAGTATCGTAAAGTTCATTTG.....CCGGGTCAAG.. 631
174 omecYsGlnSerAnSerThrMetProGLYProSerLeuGlnSerPro 191
    |||      :::::      |||      |||||      |||
632 .....GAQTACAGAGCTTACCGCGCTTCCAGCATCTGAAAGGGTT 674
191 alSerThrPro.....AlaGLY 196
    |||
675 ATTTCCAGCGCGGCGCATCTGGCTCCCGGTATACSTGCAGCGCCGCG 724
197 LysIleGLYLeuAlaValCYsTYrAsPmetArGhrProGLYleuSerLe 213
    |||||      :::::      |||      |||      |||||
725 AAATGGGGATGTTCACTCAACAGATCGCGCTGCTCAAGCTGGCG 774
213 uAlaIleuAlaGlnAlaGLYAlaGluIleLeuThTYrProSerAlaheG 230
    :::::      |||||      |||||      |||||      |
775 GGGATGGGGCTCAGGGCGCCGAGATCATCTGC.....G 809
230 LysSerIleThrGLYProAlaHISrPGLYAlleu..... 241
    |||      |||      |||      |||      :::::
810 GCGGCTACAAACAGCGCCGACCCACATCCCATGTGTCCCAACAGCAC 859
242 .....LeuArGLAlaArGLaIleGLYTh 249
    :::::      |||      :::::      |||      :::::
860 CTGACGCTTCACACATCTCTATGATGATGAGCGGGCTTATTCAGAA 909
249 rGInCYsTYrValAlaIAlaIaGlnCYsGLYArGHSISLulYSA 266
    :      :::::      ::|||      :::::      |||      ::|||
910 CGGGGCTGTGTCGCGCGCCGCGGCAAGTGGGCGATG...GAGAGAACT 956
266 rGAlaSerTYrGLYHISerMetValAlaSPTrpTrpGLYThValVal 282
    |||||      |||||      |||||      |||||      |||||
957 GCATCTGCTCGGCACCTCTTCATCTGTGGCGCGCCAGCGGGAATGTCT 1006
283 AlaArGYsSerGLu...GLYProGLYleuCYsLeuAlaArGLIleasPLe 298
    |||      :::::      |||      |||      |||      |||
1007 GCCTCTACTACGACGCGTGAAGACGAGGTATACCGCGCGCTGCACT 1056
298 uAsnTYrLeuArGLInLeuArGArGHSIleuProValPhe...GlnHIS 314
    :::::      |||||      |||||      |||||      |||
1057 CGATCGCTGCGGGAACTCGGTGACACATCTTCACCTCAAGCAGATC 1106
314 rGArPrProAsPLeuTYrGLY 320
    |||||      |||||      |||||
1107 GTCAGCCCCAGCACTAATGGT 1126

seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:US-08-294-871A-25

seq documentation block:
? Sequence 25, Application US/08294871A
? Patient No. 3624522
? GENERAL INFORMATION:
? APPLICANT: Ikenaka, Yasuhiro
? APPLICANT: Nanba, Hirokazu
? APPLICANT: Takano, Masayuki
? APPLICANT: Yajima, Kazuyoshi
? APPLICANT: Yamada, Yukio
? APPLICANT: Takahashi, Satomi
? APPLICANT: Okubo, Kazuma

```

APPLICANT: Yamada, Kazuhiko
 APPLICANT: Hiraiishi, Yoshio
 TITLE OF INVENTION: Immobilized Enzyme Preparation and
 TITLE OF INVENTION: Process for Producing D-a-Amino Acid
 NUMBER OF SEQUENCES: 70
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/294,871A
 FILING DATE: 22-AUG-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/971,758
 FILING DATE: 12-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/917,111
 FILING DATE: 07-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/211,641
 FILING DATE: 11-APR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 400848/1990
 FILING DATE: 07-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP91/01696
 FILING DATE: 06-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 407922/1990
 FILING DATE: 27-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 078840/1991
 FILING DATE: 11-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 140051/1991
 FILING DATE: 12-JUN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP92/00739
 FILING DATE: 10-JUN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 212692/1992
 FILING DATE: 10-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP93/01101
 FILING DATE: 05-AUG-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 340078/1992
 FILING DATE: 21-DEC-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Wegner, Harold C.
 REGISTRATION NUMBER: 25,258
 REFERENCE/DOCKET NUMBER: 74129/127/AOPA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1785 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 ORIGINAL SOURCE:
 STRAIN: JM109 PAD445

; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 233..1144
 US-08-294-871A-25
 alignment_scores:
 Quality: 205.00 Length: 335
 Ratio: 1.273 Gaps: 15
 Percent Similarity: 48.060 Percent Identity: 25.970
 alignment_block:
 US-09-357-675C-21 x US-08-294-871A-25 ..
 Align seg 1/1 to: US-08-294-871A-25 from: 1 to: 1785
 50 Alavalysglnvalthr.....serthrproas 59
 164 GCGATGCTTTCCACACGTTTCCCGCGCGCTGGCGGACATCACCCTAG 213
 59 pLysglnGlnAsnphelythr..... 66
 214 GAAGAGCAGAGGTTTCATGACGTCAGATGATCTTCCATGGGACAC 263
 67Cys 67
 264 AAGTCCGATCGCGCGCGGAGACGCGAAGATGCTGTCGTCCTT 313
 68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
 314 CTCGACATGTCGACGAAAGCCGAGCGCGGCGGCAATTCATTCCTT 363
 84 uProGlu.....AlaPheAspPheAlaArg.....Asp 95
 364 CCGCGAAGCTCGCGCTTACGACCTTCTCCCGCGCTGATTCACCGACG 413
 95 roAlaGluThrLeuHisLeuSerGlu..ProLeuGlyGlyLysLeuLeu 110
 414 AGCGCGAGCTGATAGGCTTCTATGAGACCGAATGCCGCGCGTGCTC 463
 111 GluGluThrThrGlnLeuAlaArgGluCysGlyLeuThrPheLeuSerLeu 127
 464 CGTCCACTTTTGAGAAAGCCCGGAACTCGGATCGGCTTCATATCGGG 513
 127 yGlyPheHis.....GluArgGlyGlnAspTrpGlnThrGln 141
 514 CTACGCTGAACCTCGTCGAAAGCGCG.....GTCAAGC 548
 141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
 549 GTCCCTTCAACACGCTCCATTTTGGTGATAGTCAGGCAGATGATGCGGC 598
 158 ThrTyrArgLysThrHisLeuCysAspValGluLeuProGlyGlnGly 174
 599 AAGTATGCTAAGATCCATTG.....CCGGGTCAACAG.. 631
 174 omeCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
 632GAGTACGAGCGCTACCGCGCTTCCACATCTTGAAGAAGCTT 674
 191 alSerThrPro.....AlaGly 196
 675 ATTTCGAGCGCGGCGATCGCTGCCGCTTCCCGGTATGACGTCGACCGCG 724
 197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
 725 AAATGGGATGTTTCATCTGCAACGATCGCGCTGCGCTGAAGCGTGGCG 774
 213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr.....TyrProSerA 228
 775 GGTATGGCGCTCAGGGCGCGGATCATCTCGGGCGGCTACACACGC 824
 228 laPheGlySerIleThrGlyProAlaHisTrpGluValLeu..... 241


```

191 alserThrPro.....Alagly 196
      |||
675 ATTTCAGAGCGGCGCATCTCGGCTCCCGGTATAGCTGACGCGCGCG 724
      |||
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerIle 213
      |||
725 AAATGGGGATGTTTCATCTCAACGATCCGCCGCTGAGCCCTGAGCC 774
      |||
213 uAlaLeuAlaGlnAlaGlyValGluIleLeuThr.....TyrProSerA 228
      |||
775 GGTCATGGGCTTCAGGGCGCGCGATCATCTCGGCGGCTTACACACGCG 824
      |||
228 laphelSerIleThrGlyProAlaHisTrpGluValLeu..... 241
      |||
825 CGACCCACATCCACCGCTCCCGACAGACACACCGAGCTGCTTCCAC 874
      |||
242 .....LeuArgAlaArgAlaIleGluThrGlnCysTyrValVa 254
      |||
875 CATCTCTATCATGATGAGCGCGGCTTATCAGAACGCGGCGCTGATCGCG 924
      |||
254 lAlaAlaAlaGlnCysGlyArgHisGlnGlyValArgAlaSerTyrGlyH 271
      |||
925 GGCCTGGGCGGAGGTGGGCTG...GAGGAGAACTGCATGCTGCTGCGGC 971
      |||
271 lSerMetValValAspProTrpGlyThrValAlaAlaArgCysSerGlu 287
      |||
972 ACTCTGTCATCATGTCGGCGCGCGGGAATGCTGCTCTCCTACAGACG 1021
      |||
288 ...GlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgI 303
      |||
1022 CTGAGAGACGAGATGATCACCGCGCGCTCATCTGATCGTCGCGGGA 1071
      |||
303 nLeuAArgArgHisLeuProValPhe...GlnHisArgArgProAspLeu 319
      |||
1072 ACTGCGTGAACACATCTCAACTCAAGACGACATCGTCAACCCAGACAT 1121
      |||
319 YrGly 320
      |||
1122 ATGGT 1126

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-294-871A-5

seq_documentation_block:

Sequence 5, Application US/08294871A

Patent No. 5824522

GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro

APPLICANT: Namba, Hirokazu

APPLICANT: Takano, Masayuki

APPLICANT: Yajima, Kazuyoshi

APPLICANT: Yamada, Yukio

APPLICANT: Takahashi, Satomi

APPLICANT: Okubo, Kazuma

APPLICANT: Yamada, Kazuhiko

APPLICANT: Hiraishi, Yoshiro

TITLE OF INVENTION: Immobilized Enzyme Preparation and

TITLE OF INVENTION: Process for Producing D-a-Amino Acid

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294,871A

```

      FILING DATE: 22-AUG-1994
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/971,758
      FILING DATE: 12-APR-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/917,111
      FILING DATE: 07-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/211,641
      FILING DATE: 11-APR-1994
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 400848/1990
      FILING DATE: 07-DEC-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP91/01696
      FILING DATE: 06-DEC-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 407922/1990
      FILING DATE: 27-DEC-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 078840/1991
      FILING DATE: 11-APR-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 140051/1991
      FILING DATE: 12-JUN-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP92/00739
      FILING DATE: 10-JUN-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 212692/1992
      FILING DATE: 10-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP93/01101
      FILING DATE: 05-AUG-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 340078/1992
      FILING DATE: 21-DEC-1992
      ATTORNEY/AGENT INFORMATION:
      NAME: Hegner, Harold C.
      REGISTRATION NUMBER: 25,258
      REFERENCE/DOCKET NUMBER: 74129/127/AOPA
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202) 672-5300
      TELEFAX: (202) 672-5399
      TELEX: 904136
      INFORMATION FOR SEQ ID NO: 5:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1785 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      ORIGINAL SOURCE:
      STRAIN: JMI09, PAD404 (FERM BP-3913)
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 233..1144
      US-08-294-871A-5

```

alignment_scores:

Quality:	Ratio:	Length:
204.00	1.267	340
Percent Similarity: 47.353	Percent Identity: 25.294	Gaps: 15

alignment_block:

US-09-357-675C-21 x US-08-294-871A-5 ..

Align seg 1/1 to: US-08-294-871A-5 from: 1 to: 1785

```

50 AlayaIcysGlnValThr.....SerThrProAs 59
      |||
164 GCGATGTGTTTCACACGTTTCCCGCGCGCTGGGCGGACATCACCTAG 213

```

59 pLysGlnGlnAsnPhenylsThr..... 66
|||:|||||:|||||
214 GAGGAGCAGAGGTTTCATGACACGTCAGTGTACTGTCAGTGGGACAC 263
67Cys 67
264 AAGTCCGATCCGCGCGGAGACACGCGAAGAGTGTGTCGTCCT 313
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
|||:|||||:|||||
314 CTCGACATGCTGACGAAAGCCGCGACCGCGCGCAATTTCATTCTCT 363
84 uProGlu.....AlaPheAspPheLeuAlaArg.....Asp 95
:|||||:|||||:|||||
364 CCCCCAACCTCGCGCTTACGACCTTCTCCCGCGCTGCGATTTCACCGCAG 413
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLeu 110
|||:|||||:|||||
414 AGCGGAGCTCGATAGCTTCTATGAGACCGAATGCCCGCGCGTGTGTC 463
111 GluGluTyrThrGlnLeuAlaArgGlyCysGlyLeuTrpLeuSerLeu 127
|||:|||||:|||||
464 CPTCCACTCTTGTAGAGAGCGCGCGACTCGGGATCGCTCAATCTGGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141
|||:|||||:|||||
514 CTACGCTGACTCTGCTGTAAGCGCGC.....GTCAAGC 548
141 ySileTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||:|||||:|||||
549 GTGCGTTCAACACGTCATTTGGTGATAGTCAGGCAAGATCGTCGCGC 598
158 ThrTyrArgLysThrHisLeuCysAspValGluLeuProGlyGlnGly 174
|||:|||||:|||||
599 AAGTATCGTAAGATCCATTTC.....CCGGTCACACAG... 631
174 oMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191
|||:|||||:|||||
632GAGTACGAGGCTTACCGCGCTTCCAGCATCTTGAAGACGCT 674
191 aSerThrPro.....AlaGly 196
|||:|||||:|||||
675 ATTTCAGCGCGCGCATCTCGGCTCCCGTATGATGATGCGCGCGG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213
|||:|||||:|||||
725 AAAATGGGATGTTCATCTGCAACGATCCGCGCTGCGCTGAAGCGTGGC 774
213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThrTyrProSerAlaPheG 230
|||:|||||:|||||
775 GGTGATGGCTCAGGCGCGCGAGTCAATCTGC..... 809
230 ySerIleThrGlyProAlaHisTyrGluValLeu..... 241
|||:|||||:|||||
810 GCGGCTACAAACACGCGACCAATCCCTGTGTCCAGACGACGAC 859
242LeuArgAlaAlaIleGluThr 249
|||:|||||:|||||
860 CTGACGCTCTTCACCATCTCTATGATGACGCGCGGTCTTATCAGAA 909
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisGlnLys 266
|||:|||||:|||||
910 GCGGCGCTGGTCCGCGCGCGCAAGTGGGCGATG...GAGGAAACT 956
266 rGalaSerTyrGlyHisSerMetValAlaAspTrpTrpGlyThrVal 282
|||:|||||:|||||
957 GCATGCTGCTGGCCACTCTGATGCTGGCGCGCGGAAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
|||:|||||:|||||
1007 GCTTCACATACGACGCTGGAAGACAGAGTGTATCACCGCGCGCTGATCT 1056

298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
|||:|||||:|||||
1057 CGATCGCTCGCGGAGACTGCTGTAACACATCTTCACTTCAACAGCATC 1106
314 rGArgProAspLeuTyrGly 320
|||:|||||:|||||
1107 GTCAGCCCGCACGACTATGCT 1126
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-294-871A-59
seq_documentation_block:
; Sequence 59, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Namba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraiishi, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; APPLICATION DATA:

```

APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 74129/127/AOPA
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109 pAD456
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-59

```

```

alignment_scores:
Quality: 204.00      Length: 340
Ratio: 1.267         Gaps: 15
Percent Similarity: 47.353      Percent Identity: 25.294

```

```

alignment_block:
US-09-357-675C-21 x US-08-294-871A-59 ..

```

```

Align seg 1/1 to: US-08-294-871A-59 from: 1 to: 1785

```

```

50 AlAAlcysGlnValThr.....SerThrProAs 59
164 GCGATGCTGTTACACAGCTTTCCGGCGCGTGGCGGACATCACTAG 213
59 pLysGlnGlnAspRheLysThr..... 66
214 GAAAGAGCAGAGTTTCATGACACAGTCAGATGATCTTCAGTGGACAAC 263
67 .....Cys 67
264 AAGTCCGATGCGCGCGGAGACAGCGGAACAGTGTCTGTCTT 313
68 AlAgluLeuValArgGluAlaAlaArgLeuGluAlaCysLeuAlaPhe 84
314 CTCGACATGCTGACGAAAGCGCGGAGCGCGGAGATTTCATGTCTT 363
84 uProGlu.....AlaPheAspRheLeuAlaArg.....Asp 95
364 CCGCGAAGTCCGCTTACGACCTTCTCCCGCGCTGATTTTACCGAGC 413
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGluGlyLeuSerLeu 110
414 AGCCCGAAGTCCGATGCTTATGACAGCGAAATCCCGCGCGGTGGTC 463
111 GluGluGluThrGlnLeuAlaArgGluCysGlyLeuThrLeuSerLeu 127
464 CGTCCAGCTTTTGAAGAGCGCGGAGACTCGGAGATCGGCTTCAATCTGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpLeuGlnThrGlnL 141
514 CTACGCTGAAGTCTGTCTGCGAAGCGGC.....GTCAAGC 548
141 ysiLeuYAspCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
549 GTCCCTTCAACAGCTGATTTTGGTAAAGTCAAGCAAGTCTGTGGC 598

```

```

158 ThrYrArgLysThrHisLeuCysAspValGluLeuProGlyGlnGlyP 174
599 AAGATTCGTAAGATCCATTGT.....CCGGGTACAGAAG... 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
632 .....GAGTACGAGGCTTACCGGCGCTTCCAGCATCTGTGAAAGCGCT 674
191 alSerThrPro.....AlaGly 196
675 ATTTCGAGCCGGCGCATCTCGGCTTCCGGTATGACGTGACGCCGCG 724
197 LysileGlyLeuAlaValCysTyraSpMetArgPheProGluLeuSerLe 213
725 AAATGGGAGATGTTTCATCTGCAAGCATGCGCGCTGCGTGAAGCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluThrTyrrProSerAlaPheG 230
775 GGTGATGGGCTTCAGGGGCGCGAGATCATCTGC.....G 809
230 LysileThrGlyProAlaHisTrpGluValLeu..... 241
810 GCGGTACACAGCGCGGACCCACATCCCTGTGTCCCGACAGCAGCAC 859
242 .....LeuArgAlaArgAlaIleGluTh 249
860 CTGACGCTCTTCCACCATCTCTATCGATGACGCGCGGTCTTATCAGA 909
249 rGlnCysTyraValAlaAlaAlaGlnCysGlyArgHisGlnLysA 266
910 CGGGGCTGTGCGCGCGCGCGGCAAGTCAGGCAATG...GAGGAGACT 956
266 rGAlaSerTyrrGlyHisSerMetValAlaAspTrpGlyThrValAl 282
957 GCATGCTGCTGCGCCATCTCTGATCGTGGCGCGGCAATTCGTC 1006
283 AlAArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
1007 GCTCATCTACGACGCTGGAAGACGAGTGATCACCGCGCGCTGAGCT 1056
298 uAsnTyrrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
1057 CGATGCTGCGCGGAACTGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106
314 rGArgProAspLeuTyrrGly 320
1107 GTCAGCCCGCACCATATGT 1126

seq_name: /cgn2.6/plodata/2/ina/6A_COMB.seq:US-08-876-398A-5
seq_documentation_block:
; Sequence 5, Application US/08876398A
; Patient No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKADA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satoshi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```

? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/876,398A
? FILING DATE: 16-JUN-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/211,641
? FILING DATE: 11-APR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: WO PCT/JP93/01101
? FILING DATE: 05-AUG-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 340078/1992
? FILING DATE: 21-DEC-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 212692/1992
? FILING DATE: 10-AUG-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Wegner, Harold C.
? REGISTRATION NUMBER: 25,258
? REFERENCE/DOCKET NUMBER: 74129/130
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 672-5300
? TELEFAX: (202) 672-5399
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1785 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: Escherichia coli
? STRAIN: JM109 PAD404 (FERM BP-3913)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: join(233..1141)
? US-08-876-398A-5

alignment_scores:
? Quality: 204.00 Length: 340
? Ratio: 1.267 Gaps: 15
? Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
US-09-357-675C-21 x US-08-876-398A-5 ..
Align seg 1/1 to: US-08-876-398A-5 from: 1 to: 1785

50 AlaValCysGlnValThr.....SerThrProAs 59
|||||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGTGTTTCACACAGCTTTTCCGCGCTGGCCGACATCACTAG 213
59 pLysGlnGlnAsnPhelysThr..... 66
|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACAGTCAGATGATCTTGAGTGGACAAC 263
67 .....Cys 67
264 AAGTCCGATCCGCGCGCGGAGACACGCGACAGTCTGCTGCTT 313
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
|||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATCTGACGAAAGCCGCGACCGGCGCAATTTCAATGCTT 363
84 uproGlu.....AlaPheAspPheLeuAlaArg.....Asp 95
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCGGACCTCGCGCTTACGACCTTCTCCGCGCTGCGCATTTACCGCAG 413
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLeu 110
|||||:|||||:|||||:|||||:|||||:|||||:
```

```

414 AGCCGAGCTCGATAGCTTCTATGACACGAAATGCCGCGCGCTGTC 463
111 GluGluThrThrGlnLeuAlaArgGluCysGlyLeuThrLeuSerLeuG 127
|||:|||||:|||||:|||||:|||||:|||||:
464 GGTCCACTCTTTGAGAGAGCCGCGAAGCTCGGATCGGTTCAATCGGG 513
127 YGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141
|||:|||||:|||||:|||||:|||||:|||||:
514 CTACGCTGAACTCGTCTCGAAGCGGC.....GTCAGC 548
141 yLysLeuAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||:|||||:|||||:|||||:|||||:|||||:
549 GTCGCTTCAACAGCTCCATTTGTTGGATGAATGACGACAGATCGTCGC 598
158 ThrTyrArgLysThrHisLeuCysAspValGlnLeuProGlyGlnGly 174
|||||:|||||:|||||:|||||:|||||:|||||:
599 AAGATCGTAAGATCCATTTG.....CCGGGTCAAG.. 631
174 oMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191
|||:|||||:|||||:|||||:|||||:|||||:
632 .....GAGTACGAGGCTTACCGCGCTTCACAGATCTTGAAGAGCTT 674
191 alSerThrPro.....AlaGly 196
|||:|||||:|||||:|||||:|||||:|||||:
675 ATTTCGACCGCGGATCTCGGCTCCCGGCTATGACGTGACGCCGCG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSer 213
|||||:|||||:|||||:|||||:|||||:|||||:
725 AAATGGGATGTTTCATCTCAACGATCGCGCGCTGAGAGCCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyLeuLeuThrTyrProSerAlaPheG 230
|||||:|||||:|||||:|||||:|||||:|||||:
775 GGTGATGGCTCAGGCGCGCCGAGATCATCTGC.....G 809
230 LysSerIleThrGlyProAlaHisTrpGluValLeu..... 241
|||||:|||||:|||||:|||||:|||||:|||||:
810 GCGGCTCAACACGCGCAGCCACATCCCTGTGTCCCGACGACGACC 859
242 .....LeuArgAlaArgAlaIleGluThr 249
|||||:|||||:|||||:|||||:|||||:|||||:
860 CTGACGCTCTCCACCATCTCTATCGATCGACGCGGCTTATACGAA 909
249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisGlnLeuSA 266
|||:|||||:|||||:|||||:|||||:|||||:
910 GCGGCGCTGCGCGCGCGCGGCAAGTGGCATC...GAGGAGACT 956
266 rGlaSerTyrGlnHisSerMetValValAspProTrpGlyThrValVal 282
|||||:|||||:|||||:|||||:|||||:|||||:
957 GCATGCTGCTCGGCGCACTCTGCATCTGGCGCGCGGGAAGATGCTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
|||:|||||:|||||:|||||:|||||:|||||:|||||:
1007 GCCTCTCACTACGACGCTGGAAGACGAGTGATGATACCCGCCCTGCATCT 1056
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
|||||:|||||:|||||:|||||:|||||:|||||:
1057 CGATCGCTCGCGGAACTGCTGAACACATCTTCACTTAAGGAGCATC 1106
314 rArgArgProAspLeuTyrGly 320
|||||:|||||:|||||:|||||:|||||:|||||:
1107 GTCAGCCCGACGACTATGCT 1126

seq_name: /cgn2.6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-59
seq_documentation_block:
? Sequence 59, Application US/08876398A
? Patent No. 6083752
? GENERAL INFORMATION:
? APPLICANT: IKENAKA, Yasuhiro
? APPLICANT: NAKABA, Hirokazu
? APPLICANT: TAKANO, Masayuki
? APPLICANT: YAJIMA, Kazuyoshi
? APPLICANT: YAMADA, Yukio
```



```

632 .....GAGTACGAGCGCTACCGCGCTTCCAGCATCTTGAAGCGGT 674
191 alserThPro.....Alagly 196
675 ATTTCGACCGGGCGATCTCGGCTTCCGGTCTATGACGTGACCGCGGG 724
197 LysileGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerle 213
725 AAAATGGGATGTTTCATCTGCAACGATCGCGCTGGCTGAAGCGTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThyTyrProSerAlaPheG 230
775 GGTGATGGGCTTCAGGGCGCGCATCATCTGC.....G 809
230 LysleThleThGlyProAlaHisTyrPgluValLeu..... 241
810 GCGGCTACACACGCGGACCCACATCCCGCTGTTCCCGACGACGACAC 859
242 .....LeuArgAlaArgAlaIleGluTh 249
860 CTGACGCTCTCCACCATCTCTCATGATGACGCGGGTCTATCACGA 909
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisGluLysA 266
910 CCGGCGCTGTCGCCGCGCGCGGCAAGTGGGATG...GAGGAGAACT 956
266 rGAlaSerTyrGlyHisSerMetValAlaAspProTyrPgluThrVal 282
957 GCATGCTCTGCGCCATCTGTCATCTGCGCGCGGCAAGGAAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
1007 GCTCATCATGACGCTGGAAGACGAGGATGATCACCGCGCGTGTGATCT 1056
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
1057 CGATGCTGCGCGGGAACGCGTGAACACATCTTCAACTTCAAGCAGCATC 1106
314 rArgProAspLeuTyrGly 320
1107 GTCAAGCCCGACACTATGCT 1126

seq_name: /cgn2.6/prodata/2/lna/6A_COMB.seq:us-08-876-398A-61

seq_documentation_block:
; Sequence 61, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: YAMAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 pMD468
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-61

```

```

alignment_scores:
  Quality: 203.00      Length: 340
  Ratio: 1.261        Gaps: 15
  Percent Similarity: 47.353      Percent Identity: 25.294

```

```

alignment_block:
US-09-357-675c-21 x US-08-876-398A-61 ..

```

```

Align seg 1/1 to: US-08-876-398A-61 from: 1 to: 1785

```

```

50 AlaValCysGlnValThr.....SerThrProAs 59
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
164 GCGATGTCTTCACAACTTTTCCGCGCGCTGGCGGACATCATCTAG 213
59 pLysGlnGlnAsnPheLysThr..... 66
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
214 GAAGGAGCAGAGGTTCTATGACACGTCAGATGATCTGACGTGGGACAAC 263
67 .....Cys 67
264 AAGTTCGATCCGCGCGGAGACACGCAAGAGTGTGCTGCTT 313
68 AlaGluLeuValArgLysAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
314 CTGCACTGCTGACGAAGCCCGAGCGGGCGGAATTTCATTGTCTT 363
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
364 CCGGACACTCGGCTTACGACCTTCTCCGGCTGGCATTTCCACCGAG 413
95 rAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
414 AGCGCGAGCTGATGATCTTCTATGACGCAAGATCCCGCGCGGTGTC 463
111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTyrPheSerLeu 127
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
464 GCTCACCTTTTGAGAGGCGCGGAACTCGGAGTCGCTCAATCTGGG 513

```

127 yelYphenis.....GluArgIysIAspTrpGluInThrGlu 141
1 111 111
514 CTACGGCTGAACTCGTCGAAAGGGGC.....GTCAGC 548
141 ysIleTrAsnCyShsValLeuLeuAsnSerIysGlyAlaValAla 157
1 111 111
549 GTCCGCTTCAACACGTCATTTTGTGATATAGTCAGGCAAGATCGCGC 598
158 ThrTrAlrLysThrHisLeuCyAspValGluIleProGlyGlnGly 174
1 111 111
599 AAGTATCGTAAGATCCATTG.....CCGGTCACAAG.. 631
174 oMetCyGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
1 111 111
632GAGTACGAGAGCTACCGCGCTTCCAGCATCTTGAAGACGTT 674
191 aIserThrPro.....AlaGly 196
1 111
675 ATTTGAGCGCGGCGATCTCGGCTCCGGTCTATGACGTCAGCGCGCG 724
197 LysIleGlyLeuAlaValCysTrpAspMetArgPheProGluLeuSer 213
1 111 111
725 AATAATGGGATGTCTATCTGCAACGATCGCGCTGCGCTGAAGCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleuThrTrpProSerAlaPhe 230
1 111 111
775 GGTGATGGGCGCTCAGGGCGCGCATCATCTGC.....G 809
230 LysIleThrGlyProAlaHisTrpGluValLeu..... 241
1 111 111
810 GCGGCTACACACGCGCCACCAATCCGCTGTCCCGACGACGACAC 859
242LeuArgAlaArgAlaIleGluThr 249
1 111 111
860 CTGACGCTCTCCACCATCTCTATGATGAGGCGCGGCTTATCAGAA 909
249 rGlnCysTrpValValAlaAlaIleGlnCysGlyArgHisIleGluLys 266
1 111 111
910 CGGGGCGCTGTCGGCGCGCGCGCAAGGGGCGCATC..GAGGAGACT 956
266 rAlaSerTrpGlyHisSerMetValValAspProTrpGlyThrVal 282
1 111 111
957 GCATGCTGCTCGGCGCATCTGATCGCGCGCGCAAGGGAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCySLeuAlaIleAsp 298
1 111 111
1007 GCTCTACTACGACGCTGGAAGACGAGGATCATCCGCGCGCTGATCT 1056
298 uAsnTrpLeuArgGlnLeuArgHisLeuProValPhe...GlnHis 314
1 111 111
1057 CGATCGCTCCCGGAGACTCGGTAAACATCTTCAACTTCAACGAGCATC 1106
314 rGArgProAspLeuTrpGly 320
1 111 111
1107 GTACGCGCCACGACATATGCT 1126

seq_name: /cgn2_6/ptodata/2/ine/5A_COMB.seq:US-08-294-871A-9

seq_documentation_block:
: Sequence 9, Application US/08294871A
: Patent No. 5824522
: GENERAL INFORMATION:
: APPLICANT: Ikenaka, Yasuhiro
: APPLICANT: Nanda, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hirataishi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and

TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109 PAD416 (FERM BP-3915)
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144

US-08-294-871A-9

alignment_scores:
 Quality: 202.00 Length: 340
 Ratio: 1.255 Gaps: 15
 Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
 US-09-357-675C-21 x US-08-294-871A-9 ..

Align seg 1/1 to: US-08-294-871A-9 from: 1 to: 1785

```

50 AlaValCysGlnValThr.....SerThrProAs 59
164 GCGATGTGTTTCACAAACGTTTCCCGCGCGTGGCGCGACATCCTAG 213
59 pLysGlnGlnAsnPhelYsThr..... 66
214 GAAGGAGCAGAGGTTTCATGACACGTCAGATGACTTGCAGTGGACAAC 263
67 .....Cys 67
264 AAGTCCGATCGCGCGCGGAGACACGCGAAGAGTGTCTGCTGCTT 313
68 AlaGlnLeuValArgLysAlaAlaArgLeuGlnAlaCysLeuAlaPheLe 84
314 CTCGACATGCTGACGAAGCGCGGAGCGCGGCGGAATTTCATCTCTT 363
84 uProGlu.....AlaPheAspPheLeAlaArg.....Asp 95
364 CCCGCAACTCGCGCTTACGACCTTCTCCCGCGCTGGCATTTACCGACG 413
95 roAlaGlnThrLeuHisLeuSerGlu..ProLeuGlnGlyLysLeu 110
414 AGCGCGAGCTCGATAGCTTCTATGAGACCGGAATGCGCGCGGCGG 463
111 GlnGluTyrThrGlnLeuAlaArgLysGlnGlyLeuTrpLeuSerLeu 127
464 CGTCACTCTTTGAGAGAGCGCGGACACTCGGATCGGCTTCACTGG 513
127 yGlyPheHis.....GluArgGlnGlnAspTrpGlnGln 141
514 CTACGCTGAACTCGTCTGCAAGGCGCG.....GTCACAC 548
141 ySileTyrAsnGlyHisValLeuLeuAsnSerLysGlnValAlaValAla 157
549 GTCCGCTTCAACACGTCATTTGGTGATAGTCAAGCAAGTCTGCGC 598
158 ThrTyrArgLysThrHisLeuCysAspValGlnLeuProGlnGln 174
599 AAGTATCGTAAGATCCATTG.....CCGGTCACACAG.. 631
174 cMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191
632 .....GAGTACGAGGCTTACCGCGCTTCCAGCATCTTGAAGCGCT 674
191 alSerThrPro.....AlaGly 196
675 ATTTCGAGCGCGCGCATCTCGGCTTCCCGGTATGACGTCACGCGCG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGlnLeuSer 213
725 AAAATGGGATGTTCATCTGCAACGATCGCGCTGGCCGGAAGCCGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGlnLeuThrTyrProSerAlaPheG 230
775 GGTGATGGGCTTCAGGCGCGCGAGATCATCTGC.....G 809
230 LysLeuThrGlyProAla..... 236
810 GCGGCTACACACGCGCGACCAATCCCTGTTCCCGACGACGACAC 859

```

```

237 .....HisTrpGlnValLeuLeuArgAlaAlaIleGln 249
860 CTGACGCTCTTCACCATCTCCTATGATGACAGCGCGCTTATCAGAA 909
249 rGlnCysTyrValAlaAlaAlaGlnGlnCysGlyArgHisGlnGly 266
910 CGGGGCTGTGCGCGCGCGCGGACAGGCGGCGATG..GAGGAAACT 956
266 rGlnAsnSerTyrGlnHisSerMetValAlaAspTrpGlnThrVal 282
957 GCATGCTGTGCGCGCATCTCTCATGCTGCGCGCGCGGGAATCTG 1006
1007 GCTCTCAGTACGACGCTGGAACGAGGATGATCACCAGCGCGCTGAT 1056
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHis 314
1057 CGATCGCTGCGCGGACGCTGTAACATCTTCAACTTCAACGACGATC 1106
314 rArgProAspLeuTyrGly 320
1107 CTCAGCGCCACGACTATGCT 1126

seq_name: /cgn2_6/p/tdaata/2/1na/5A_COMB.seq:US-08-294-871A-11
seq_documentation_block:
: Sequence 11, Application US/08294871A
: Patent No. 5824522
: GENERAL INFORMATION:
: APPLICANT: Ikenaka, Yasuhiro
: APPLICANT: Nanda, Hirokazu
: APPLICANT: Yajima, Masayuki
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hirashi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and
: NUMBER OF SEQUENCES: 70
: CORRESPONDENCE ADDRESS:
: ADDRESS: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/294,871A
: FILING DATE: 22-AUG-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,758
: FILING DATE: 12-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/917,111
: FILING DATE: 07-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/211,641
: FILING DATE: 11-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 400848/1990
: FILING DATE: 07-DEC-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP91/01696
: FILING DATE: 06-DEC-1991

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold G.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109 PAD428
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-11

alignment_scores:
Quality: 202.00 Length: 340
Ratio: 1.255 Gaps: 15
Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
US-09-357-675c-21 x US-08-294-871A-11 ..
Align seg 1/1 to: US-08-294-871A-11 from: 1 to: 1785

50 AlalaIcysGlnValThr.....SerThrProAs 59
|||||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGTGTTTCAACAACGTTTCCGCGCGCTGGCGGAGACATCAGTAC 213
59 pIySgInGlnAsnPhelYsThr..... 66
|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACAGTCAGATGATCTTGCAGTGGACAAC 263
67 .....Cys 67

264 AAGTCCGATCCGCGCGCGGAGACACGGAACAGGTCGTCGTCGTT 313
68 AlagIuLeuValArgIuAlaIaArGLyAlaCysLeuAlaPheLe 84
|||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATCTGACGAAGCCGCGAGCCGGCGCAATTTCATTGTCTT 363
84 uDrOglu.....AlaPheAspPhelLeAlaArg.....Asp 95
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCCCAGAACTCGCGCTTACGACCTTCTTCCCGCGCTGCATTTACCGAG 413
```

```

95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLeu 110
|||||:|||||:|||||:|||||:|||||:|||||:
414 AGGCCAGCTCCGATAGCTTCTATGAGACGGAATGCCGCGCGGTGTC 463
111 GluGluTyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 127
|||||:|||||:|||||:|||||:|||||:|||||:
464 CGTCACCTCTTGAAGAAGCCGCGGAACCTCGGATCGCTCAATCGGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141
|||||:|||||:|||||:|||||:|||||:|||||:
514 CTCAGCGTGAACCTGCTGTCGAGAGCGGC.....GTCAGC 548
141 ySLeTyraSncysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||||:|||||:|||||:|||||:|||||:|||||:
549 GTCCGCTTCAACAGCTCATTTTGTGATGATGACGAGCAAGATCGTGC 598
158 ThrTyraGlyThrHisLeuCysAspValGluIleProGlyGlnGly 174
|||||:|||||:|||||:|||||:|||||:|||||:
599 AAGTATCGTAAATCCATTG.....CCGGTCAACAG.. 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
|||||:|||||:|||||:|||||:|||||:|||||:
632 .....GAGTACGAGCGCTACCGCGCTTCCAGATCTTGAAGAAGCTT 674
191 alSerThrPro.....AlaGly 196
|||||:|||||:|||||:|||||:|||||:|||||:
675 ATTTGAGCCGCGGCGATCTCGGCTTCCGCTATGACGTGACGCCGCG 724
197 LysIleGlyLeuAlaValCysTyraSpmetArgPheProGluLeuSerLe 213
|||||:|||||:|||||:|||||:|||||:|||||:
725 AAATGGGATGTTTCATCTGCAACAGATCGCGCTGAGCTGAAGCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyraProSerAlaPhe 230
|||||:|||||:|||||:|||||:|||||:|||||:
775 GGTGATGGCGCTCAGGGCGCGAGATCAATCTGC.....G 809
230 LysSerIleThrGlyProAla..... 236
|||||:|||||:|||||:|||||:|||||:|||||:
810 GCGGCTACACAGCCGACCCACAAATCCCTGTCCCGACGACGAC 859
237 .....HisTrpGluValLeuLeuArgAlaIleGluThr 249
|||||:|||||:|||||:|||||:|||||:|||||:
860 CTGACGCTCTCCACCATCTCTATGATGACGCGCGGTATATCGAA 909
249 rGlnCysTyraValAlaAlaAlaGlnCysGlyAlaGlnHisGluLysA 266
|||||:|||||:|||||:|||||:|||||:|||||:
910 CGGGCGCTGTCGCGCGCGCGGCAAGCGGCGCATG...GAGGAGAACT 956
266 rGlnAspSerTyrglyHisSerMetValAlaAspProTrpGlyThrVal 282
|||||:|||||:|||||:|||||:|||||:|||||:
957 GCATGCTGCTGGCCACTCTGATCGTGGCGCGACCGGGAATGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
|||||:|||||:|||||:|||||:|||||:|||||:
1007 GCCTCCTACATGACAGCGTGAAGAGAGAGGTATGACCGCGCGCGATCT 1056
298 uAsnTyraLeuArgGlnLeuAlaArgHisLeuProValPhe...GlnHis 314
|||||:|||||:|||||:|||||:|||||:|||||:
1057 CGATCGCTCGCGGGAACCTCGTGAACACATCTTCAACTTCAGACGATC 1106
314 rGlnArgProAspLeuTyrgly 320
|||||:|||||:|||||:|||||:|||||:|||||:
1107 GTACGCCCGACACTATAGT 1126

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-294-871A-21
seq_documentation_block:
; Sequence 21, Application US/08294871A
; Patent No. 582452
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Namba, Hirokazu
; APPLICANT: Takano, Masayuki
```

```

APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okudo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiratsuki, Yoshio
TITLE OF INVENTION: Immobilized Enzyme Preparation and
NUMBER OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
City: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/J91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078640/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/J92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/J93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SRO ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
type: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; STRAIN: JM109 PAD439
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 233..1144
US-08-294-871A-21

alignment_scores:
    quality: 202.00      Length: 340
    Ratio: 1.255         Gaps: 15
    Percent Similarity: 47.353      Percent Identity: 25.294

alignment block:
US-09-357-675C-21 x US-08-294-871A-21 ..

Align seg 1/1 to: US-08-294-871A-21 from: 1 to: 1785

50 AAlaValcysGlnValThr.....SerThrProAs 59
|||||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGTCGTTTCACACACTTTTCCCGGCGCGGGCCGAGACATCACTG 213
59 pLVSGlGlnAsnRheIyThr..... 66
|||||:|||||:|||||:|||||:|||||:
214 GAAGGACGACAGAGTTTCATGACACGTCAGATGATCTTGCGATGGGACAA 263
67 ..... 67
264 AAGTCCGATCGCGCGCGGAGACACCGCGAACAAGTCGTCGTTGCTTT 313
68 AAGlGlnLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaRhe 84
|||||:|||||:|||||:|||||:|||||:
314 CTCGACATGCTGACGAAAGCCGCGAGCGCGGCGCGAATTTTCATGTCT 363
84 uPTGGLN.....AlaRheAspRheIleAlaArg.....Asp 95
|||||:|||||:|||||:|||||:|||||:
364 CCGCGAAGTCGCGCTTACGACCTTCTTCCGCGCTGGCGCATTTCCACCG 413
95 roAlaGlnThrLeuHisLeuSerGlu...ProLeuGlyLysLeuLeu 110
|||||:|||||:|||||:|||||:|||||:
414 AAGCGAGAGTCGATAGCTTTATGAGACCGAAGTCCGCGCGGCGGTC 463
111 GlnGluTyThrGlnLeuAlaArgGlyCysGlyLeuThrPLeuSerLeu 127
|||||:|||||:|||||:|||||:|||||:
464 CGTCCACTCTTGGAGAGCGCGCGAATTCGCGATCGGCTCAATCTGGG 513
127 yGluRheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141
|||||:|||||:|||||:|||||:|||||:
514 CTACGCTGAACCTGCTCGTAAGCGGC.....GTCAAC 548
141 ySileTyAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||||:|||||:|||||:|||||:|||||:
549 GTGCGCTTCAACACGCTCATTTTGGTGAGTAAGTCAGGCAAGATCGCGC 598
158 ThrTyArgLysThrHisLeuCysAspValGluLeuProGlnGlyR 174
|||||:|||||:|||||:|||||:|||||:
599 AAGATGCTAAGATCAATTTG.....CCGGTCACAG.. 631
174 oMetCysGlnSerAsnSerThrMetProGlyProSerLeuGlnSerPro 191
|||||:|||||:|||||:|||||:|||||:
632 .....GAGTACGAGGCGCTACCGCGCTTCACAGATCTTGAAGCGTT 674
191 AlSerThrPro.....AlaGly 196
|||||:|||||:|||||:|||||:
675 ATTTCGACCGCGGCGATTCGCGCTTCCGCGTATGACGTCGACGCGCG 724
197 LysIleGlyLeuAlaValCysTyTrpMetArgPheProGlnLeuSer 213
|||||:|||||:|||||:|||||:|||||:
725 AAAATGGGATGTCATCTGCAACGATGCGCGCTGGCTGCGAAAGCGGG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAlaPheG 230

```



```
775 GGGATGGGCTCAGGGGCGGAGATCATCTGC.....G 809
230 YSerIethrclYProAla..... 236
810 GGGGTACACAGCGCGACCAATCCCCGTGTCCCGACGACGACC 859
237 .....HisTrpGluValLeuArgAlaArgAlaIleGluHr 249
860 CTGACGCTCTTCACCATCTCTATGATGACAGCGCGGTCTTATCAGA 909
249 rGlnCysTrValAlaAlaAlaAlaIncysglYArGHisHISgluLysA 266
910 CCGGGCGCTGTCGGCGCGCGGCAAGGTGCGATG...GAGGAAAC 956
266 rGAlaserTrglYHisSerMetValAlaAspTrpTrpGlyHrVal 282
957 GCATGCTGCTCGCGCATCTGATGTCGCGCGGACCGGGAATCGTC 1006
283 AlArGcYsSerGlu...GlyProGlyLeuCysLeuAlaArgIleAsp 298
1007 GCTCTACTACGACGCTGAGACGAGGTGATCACCAGCGCGCTCATCT 1056
298 uAsnTrpLeuArgGlnLeuArgHISleuProValPhe...GlnHISa 314
1057 CGATCGCTCGCGGAGACTGCTGACACACATCTTCACTTCAACGACATC 1106
314 rGArGProAspLeuTrgIy 320
1107 GTCAGCGCCGACGACATATGCT 1126
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-9

```
seq_documentation_block:
Sequence 9, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NANBA, Hirokazu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESS: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
```

```
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD416 (FERM BP-3915)
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-9
```

```
alignment_scores:
Quality: 202.00 Length: 340
Ratio: 1.255 Gaps: 15
Percent Similarity: 47.353 Percent Identity: 25.294
```

alignment_block:
US-09-357-675c-21 x US-08-876-398A-9 ..

Align seg 1/1 to: US-08-876-398A-9 from: 1 to: 1785

```
50 AlAlaValcysGlnValThr.....SerThrProAs 59
164 GCGATGTTGTTACAAACGTTTCCCGCGCTGGCGGACATCACCTAG 213
59 pLysGlnGlnAsnPhelLysThr..... 66
214 GAAGAGACAGAGGTTCTATGACAGTACATGATGATACATTGACATGGACAC 263
67 .....Cys 67
264 AAGTCCGATCGCGCGCGGAGACACGCAACAGTCTGCTGCTT 313
68 AlAGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 84
314 CTGACATGCTGACGAAGCGCGGAGCGCGGATTTCAATTGCTT 363
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
364 CCCGAACTCGCGCTTACGACCTTCTCCGCGCTGACATTCACCGACG 413
95 roAlAGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLeu 110
414 AGCCGAGCTCGATAGCTTATGAGACGGAATCCCGCGGTGTC 463
111 GluGluTrpThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG 127
464 CGTCCACTCTTGAAGAAGCGCGGAACTGGGATCGGCTTCAATCGGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGln 141
514 CTACGCTGACCTGCTCGTCAAGCGCGC.....GTCACG 548
141 ySleTrpAsnGlyHisValLeuLeuAsnSerLysGlyAlaValAla 157
549 GTCGCTTCAACAGTCCTCATTTGTGGATGATGACGCAAGATCGCGG 598
158 ThrTrpArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 174
599 AAGTATGTAAGATCATTTG.....CCGGTCAACAAG.. 631
```



```
464 CGTCCACTCTTTGAGAAGCGCGGAACTCGGATCGGCTTCAATCTGGG 513
127 yGIpPheHis.....GLuArgGLyGlnAspTIpGLuGlnThrGlnL 141
514 CTACGCTGAAGCTCGTCGTCGAGCGGC.....GTCAGC 548
141 ysIleTyraNcysHisValleuLeuAsnSerLysGlnAlaValAla 157
549 GTGCGTTCAACAGTCATTTTGTGGATAGTCAGGCAAGATCGTCGCC 598
158 ThrTyArgLysThrHisLeuCysAspValGluIleProGlnGlnLPr 174
599 AAGTATCGTAAGATCCATTG.....CCGGTCCACAG.. 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
632 .....GAGTACGAGCGCTACCGCGCTTCACGATCTTGAAGCGCT 674
191 aIserThrPro.....AlaGly 196
675 ATTTCCAGCCGGCGATCTGGCTCCCGGTATGACGTGAGCGCG 724
197 LysIleGlyLeuAlaValCysTyraSpmetArgPheProGluLeuSerLe 213
725 AAATGGGGATGTTCACTGCAACGATCGCGCTGCGCTGAGCGCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAlaPheG 230
775 GGTGATGGGCGCTCAGGGCGCGCATCTGC.....G 809
230 IySerIleThrGlyProAla..... 236
810 GCGGCTACACACGCGACCCACAAATCCCTGTCCCAAGACGACCAC 859
237 .....HisTrpGluValLeuLeuArgAlaIleGluThr 249
860 CTGACGTCCTCCACCATCTCTATGATGACAGCGCGGTCTTATCAGAA 909
249 rGlnCysTyrValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
910 CGGGGCTGTGGTCCGCGCGCGGCAAGCGCGCATG..GAGGAGAACT 956
266 rGlnAserTyrgLysSerMetValAlaAspProTrpGlyThrValVal 282
957 GCATGCTGCTCGGCACCTCTGCACTGGCGCGACCGGGGAAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
1007 GCTCTCACTACGACGCTGGAAGAGAGTATCATCCGCCGCTGATCT 1056
298 uAsnTyrLeuArgGlnLeuArgHisLeuProValPhe...GlnHisA 314
1057 CGATCGCTGCGGGAACTCGGTGAACATCTTCAACTTCAAGCAGCATC 1106
314 rGArgProAspLeuTyrgLy 320
1107 GTCAAGCCCAAGCACTATGTT 1126
```

THIS PAGE BLANK (USPTO)

OM of: US-09-357-675C-22 to: GenBank: * out_format : pfs

Date: Apr 29, 2002 11:10 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+2n.model -DEV=x1h  
-O=/gen2.1/USPRO-pool/US09357675/runat_29042002_091757_3179/app-query.fasta_1.1208  
-DB=genbank1 -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELTEXT=7.000 -START=1 -MATRIX=bloms62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=ptc  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09357675 @CGN1_1 10065 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPX -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-357-675C-22  
Query length: 323  
Database: GenBank: *  
Database sequences: 1797656  
Database length: 1873333701  
Search time (Sec): 3374.980000
```

score list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_ro:AF069988	1731.00	2473.72	2.0e-129	1338	AF069988 Mus musculus nitrilase
gb_ro:BC021634	1723.00	2462.07	8.9e-129	1365	BC021634 Mus musculus, nitrilase
gb_pr:AF069987	1487.00	2123.28	6.6e-110	1385	AF069987 Homo sapiens nitrilase
gb_ro:AF069985	1347.00	1912.15	3.8e-98	4481	AF069985 Mus musculus nitrilase
gb_hlg:AC084821	1343.00	1872.83	5.9e-96	211771	AC084821 Mus musculus chrom
gb_hlg:AC087229	1320.00	1840.49	3.7e-94	213203	AC087229 Mus musculus chrom
gb_pr:AF069994	1211.50	1718.52	2.2e-87	4079	AF069994 Homo sapiens chrom
gb_hlg:AL354714.4	1211.50	1689.83	9.2e-86	110000	Continuation (5 of 7) of AL3
gb_hlg:AL354714.5	1211.50	1689.83	9.2e-86	110000	Continuation (6 of 7) of AL3
gb_hlg:AL590651	1211.50	1684.59	1.5e-85	167863	AL590651 Homo sapiens chrom
gb_pr:AL591806	1211.50	1684.59	1.5e-85	200822	AL591806 Human DNA sequence
gb_hlg:AC105589	1187.00	1650.46	1.4e-83	178418	AC105589 Rattus norvegicus c
gb_ov:AF284575	933.50	1330.14	1.0e-65	1214	AF284575 Xenopus laevis N1 pr
gb_ov:AF284576	706.00	1002.53	1.8e-47	1385	AF069986 Caenorhabditis elegans
gb_hlg:AF069989	699.50	992.38	6.5e-47	1521	AF069989 Drosophila melanogast
gb_hlg:AC017755	699.50	953.53	9.5e-45	131853	AC017755 Drosophila melanogast
gb_hlg:AC093121	699.50	951.40	1.2e-44	168250	AC093121 Drosophila melanogast
gb_hlg:AC093501	699.50	951.14	1.3e-44	173494	AC093501 Drosophila melanogast
gb_hlg:AC006169	699.50	951.10	1.3e-44	174279	AC006169 Drosophila melanogast
gb_hlg:AC105293	699.50	950.33	1.4e-44	190351	AC105293 Drosophila melanogast
gb_hlg:AF284571	699.50	946.41	2.4e-44	298640	AF284571 Drosophila melanogast
gb_hlg:AF372904	699.50	902.81	6.3e-42	11172	AF372904 Arabidopsis thaliana
gb_hlg:AC010013	624.00	856.98	2.3e-39	340131	AC010013 Drosophila melanogast
gb_hlg:AE000934	538.50	741.34	6.3e-33	15124	AE000934 Methanobacterium th
gb_pl:SPBC651	529.00	723.21	6.4e-32	25350	AL033570 S.pombe chromosome I
gb_pl:SCY0116W	524.00	741.86	5.9e-33	1307	249401 S.cerevisiae chromosome
gb_pl:AF284571	524.00	740.14	7.3e-33	1591	AF284571 Saccharomyces cerevis
gb_pl:SCGCD14	497.00	696.50	2.0e-30	27912	27912 S.cerevisiae GC014 gene
gb_ba:SYGSLRD	497.00	657.27	9.0e-28	110908	D64002 Synchocytis sp. PC
gb_ba:AB017194	491.50	683.96	3.8e-30	4762	AB017194 Plectonoma botryum C
gb_ba:AE004861	488.00	672.08	4.2e-29	10466	AE004861 Plectonoma botryum C
gb_ba:AE004862	483.00	685.69	7.9e-30	962	AF284574 Homo sapiens N1t prot
gb_ba:AE004863	483.00	685.69	8.4e-30	1019	AF284574 Homo sapiens N1t prot
gb_ba:AE004864	483.00	685.69	8.4e-30	965	BC020620 Homo sapiens, N1t prot
gb_ba:AF005512	479.50	668.82	6.9e-29	3180	AF005512 Nostoc commune tRNA-L
gb_ba:AP003587	478.50	628.05	1.3e-26	34350	AP003587 Nostoc commune tRNA-L
gb_ba:BC020153	478.50	628.05	1.3e-26	1063	BC020153 Mus musculus, N1t prot
gb_ba:AF284573	474.50	672.62	4.2e-29	1292	AF284573 Mus musculus, N1t prot
gb_ba:AL646071	462.00	608.95	1.5e-25	203050	AL646071 Ralstonia solanaceae
gb_ba:G72919	462.00	608.95	1.5e-25	847	G72919 MARC 2849-2850:991933517

```
seq_name: gb_ro:AF069988  
seq_documentation_block:  
LOCUS AF069988 1338 bp mRNA linear ROD 23-JUL-1998  
DEFINITION Mus musculus nitrilase 1 (N1t1) mRNA, complete cds.  
ACCESSION AF069988  
VERSION AF069988.1 GI:3228667  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 1338)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rotman,J.H., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Nitrilase and fhit homologs are encoded as fusion proteins in  
drosophila melanogaster and caenorhabditis elegans  
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
9837986  
MEDLINE  
REFERENCE 2 (bases 1 to 1338)  
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rotman,J., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Direct Submission  
Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA  
FEATURES  
source  
location/Qualifiers  
1..1338  
/organism="Mus musculus"  
/db_xref="taxon:10090"  
/chromosome="1"  
/map="1q21-q23"  
1..1338  
/gene="N1t1"  
58..1029  
/gene="N1t1"  
/codon_start=1  
/product="nitrilase 1"  
/protein_id="AAC40185.1"  
/db_xref="GI:3228668"  
/translation="MGFTIRPPHQLCTGYRLRPVICTQPRPRMTSSSTSMELPL  
VAVCOVTSFKNQENKTCALVEQEARAGACAPLEAFDFIARNDAEFLILSEPLN  
GDLGGYLSAARCGITMISGSHERGQDEONOKYINCIVILNSGSVAASRKHIL  
CYETIGGQGMRSNTTKRGTELEPPYKTRAGVGAICIDMRFPELSKLAAGAEI  
LTVPSAFSGSTGPAHWEVLIRAKAIESQCVIAAOCGRHHEHRASTGISMVVDPMGT  
VVARCSGPELCLARIDLHFLQMRHLPVFOHRRDPLGSLHPLS"  
BASE COUNT 347 a 335 c 330 g 326 t  
ORIGIN  
alignment_scores:  
Quality: 1731.00 Length: 323  
Ratio: 5.376 Gaps: 0  
Percent Similarity: 99.690 Percent Identity: 99.690  
alignment_block:  
US-09-357-675C-22 x AF069988 ..  
Align seg 1/1 to: AF069988 from: 1 to: 1338  
1 MetLeuGlyPheIleThraPrpProHnsgInleuLeuGysThrGlyTy 17  
|||||  
58 ATGCGGGGCTTCATACCAAGGCTCTCTCCACACCTGTGTGACGGGATA 107  
|||||  
17 TargLeuTauArgIleProValLeuGysThrGlnProArgProArgThrm 34  
|||||
```

```

108 CCGATTACTTGGAAACCCAGTACTTGTACTAGCCAGCCAGACCA 157
34  eISerSerSerThrSerTrpGluLeuProLeuValAlaValCysGluVal 50
158 TGTCTTATCACTTCTCTGGAGAGCTGCCCTGGTGGCTGTGTCCAGTA 207
51  ThrSerThrProAsnLysGlnGluAsnPhelSerThrcysAlaGluLeuVa 67
208 ACATCAACACCAACAGCAAGAGACTTTAAACATGTGCTGAGTTGGT 257
67  IGlnGluAlaAlaIarGluGlyAlaCysLeuAlaPheLeuProGluAlaP 84
258 TCAAGAGAGCTGCCAGACTGGGTGCTTGCCTTCTTGCCTCAGGAT 307
84  heaSPheileAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100
308 TTGACTTTATTGACAGAAACCCCTGCCAGACATTACTCTCTCCGAAACA 357
101 LeuAsnGlyAspLeuLeuGlyGlnTrpSerGlnLeuAlaArgGluCysG 117
358 CTGAATGGGGATCTTTGGCCATATAGCCAGCTTGGCAGGAAATGG 407
117 yLIETrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGlu 134
403 AATCTGGCTGTCTTGGGGCGGTTTCCAGAGCGTGGCCAAAGACTGGAGC 457
134  lnaSngLysLysIleTrpAsnGlyHisValLeuLeuAsnSerLysGlySer 150
453 AGAATCGAAATATCTACAAATGTCAATGCTTGTGAACAGCAAGGAGTCA 507
151 ValValAlaSerTrpArgLysThrHisLeuCysAspValGluIleProG 167
508 GTACTGGCCAGTTACAGGAAGACACATCTGTGCATGTAGAGATCCAGG 557
167 yGlnGlyProMetArgGluSerAsnTrpThrLysProGlyGlyThrLeu 184
558 TCAGAGGGCCGATGTAGAGAAAGCAACTATACCAAGCTGGAGGACTTGG 607
184  lUpProValLysThrProAlaGlyLysValGlyLeuAlaIleCysTrp 200
608 AGCCACCTGTCAAGACACCGGCTGGCAAGGTGGTGTACGACATCTGTAT 657
201 AspMetArgPheProGluLeuSerLeuLysLeuAlaGlnAlaGlyAlaG 217
658 GACATGGGGTCCCTGGAACCTTTTGAATTTGGCTCAGCTGGGGGAGA 707
217 uLIELeuThrTrpProSerAlaPheGlySerValThrGlyProAlaHis 234
708 AATACTTACTTATCTTCAGCCCTTGGATGTGTACAGGTCCAGCCCACT 757
234  rPgIuValLeuLeuAlaArgAlaIleGluSerGlnCysTrpValIle 250
758 GGGAGGTGCTGCTCGGGCCGCCCATGATCTCAGTGTATGTATTA 807
251 AlaAlaIaGlnCysGlyArgHisIleGluThrArgAlaSerTrpGlyHis 267
808 GCACAGCGGCGAGTGTGAGCGCCACCATGAAACAGAGCAAGATTATGGCA 857
267 sSerMetValValAspProTrpGlyThrValValAlaArgCysSerGlu 284
858 TAGCATGTGTGTGACCCGTGGGACAGTGTGCTGCCGCTGCTCGGAG 907
284  lYpProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuGlnGlnMet 300
908 GACCAAGGCTCTGCTTGGCTGGAATGTATCTCCACTTTACAAACAGATG 957
301 ArgGlnHisLeuProValPheGlnHisArgArgProAspLeuTrpGlySe 317
958 CGCAACACACGCGCTGTGTTCACAGCAGCGACCTGACCTATGGAG 1007
317  rLeuGlyHisProLeuSer 323
1008 TCTGGGTATCATCCACTTCT 1026

```

```

seq_name: gb_0:BC021634
seq_documentation block:
LOCUS BC021634 1365 bp mRNA linear ROD 22-JAN-2002
DEFINITION Mus musculus, nitrlase 1, clone MGC:13825 IMAGE:4008543, mRNA,
complete cds.
ACCESSION BC021634
VERSION BC021634.1 GI:18204912
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 1365)
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbioology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Keltman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 18 Row: e Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6754855.
FEATURES
source
Location/Qualifiers
1..1365
/organism="Mus musculus"
/db_xref="LOCUSID:27045"
/db_xref="taxon:10090"
/clone="MGC:13825 IMAGE:4008543"
/tissue.type="Mammary tumor metastasized to lung.
MMTV-LTR/Mnt1 model. Expression driven by an MMTV-LTR
enhancer."
/clone.lib="NCI-CGAP_Lu30"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
59..1030
/codon_start=1
/product="nitrlase 1"
/protein_id="AAH21634.1"
/db_xref="GI:18204913"
/translation="MLGFTTRPQHLCTGYRLRTPVLCTQPRPTSSSYSWELPL
VAVCVSTPNKQENKTCALVQEARGLAFLEPAFDEIARNPAETLLSEPLN
GDLGOYSQIAREGCIWLSIGFHERGDMRONKIYNQHVILNKGSVVAYSRKTIL
CDVEITPGGPRAESNTKPGGTLPEPRPVKTRPAKVGALCYDRFPBLSIKLQAQAEI
LTVSSAFGSVGTPAHMEVILARRALIESQCYVIAAQCGRHHTETRASTYSGSMVDPWGT
VVAARSEGGVLAIDILHFIQDMQHLPVFQHRRPDLXSGHBLIS"
BASE COUNT 368 a 334 c 331 g 332 t
ORIGIN
alignment_scores:
Quality: 1723.00 Length: 323
Ratio: 5.368 Gaps: 0
Percent Similarity: 99.381 Percent Identity: 99.381
alignment_block:
US-09-357-675C-22 x BC021634

```

Align seg 1/1 to: BC021634 from: 1 to: 1365

1 MetLeuGlyPheIleThrArgProHisGlnLeuLeuGly 17
|||||
59 ATGCTGGGGCTTCATCACCAGGCTCTCCACCACCTCGTGTACCGGTA 108
17 rArgLeuLeuArgIleProValLeuGlySerHisGlnProArgProArgThr 34
|||||
109 CCGATTACTTGAACCCAGTACTTGTACTACACCGCCAGCGCCAGAACCA 158
34 eSerSerSerThrSerTrpGlnLeuProLeuValAlaValCysGlnVal 50
|||||
159 TGTCTCATCAACTTCTGGGAGCTGCCCTGGGCTGTGTCCAGGTA 208
51 ThrSerThrProAsnLysGlnGlnAsnPheLysThrCysArgIleVal 67
|||||
209 ACATCAACACCAACAGACAGAGACTTTAAACATGTGTGAGTGGT 258
67 LglnGlnAlaAlaArgLeuGlnAlaCysLeuAlaPheLeuProGlnVal 84
|||||
259 TCAAGAGGCTGCCAGACTGGGTGCTTGCTTGCTTGCTTGCTGAGGAT 308
84 heaSpheIleAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100
|||||
309 TTGACTTTATTTGACGAAACCTGCGAGACATTACTCTGCCAACA 358
101 LeuAsnGlyAspLeuLeuGlnGlnTrpSerGlnLeuAlaArgGluCysGln 117
|||||
359 CTGATGGGATCTTTGGGCCAATATAGCCAGCTTGCAGGAAATGGG 408
117 yLeuTrpLeuSerLeuGlnGlnPheHisGlnArgGlnAspTrpGln 134
|||||
409 AATGTGGCTCTCTTGCGGCGTTCCAGACGCGTGGCCAGACAGCGGAGC 458
134 LAsnGlnLysIleTrpAsnCysHisValLeuLeuAsnSerLysGlnSer 150
|||||
459 AGAATCAGAAATCTACAAATGTGATGCTTGTGAACAGACAGGATCA 508
151 ValValAlaSerTrpArgLysTrpHisLeuGlyAspValGluIleProGln 167
|||||
509 GTAGTGGCCAGTTACAGAAACACATCTGTGCATGTAGAGATCCACGG 558
167 yGlnGlyProMetArgGlnSerAsnTrpThrLysProGlnGlyThrLeuGln 184
|||||
559 TCAGGGGCCATGAGAGAACCACTATACCAACCTCGGAGGACACTCTG 608
184 LProProValLysTrpProAlaGlyLysValGlyLeuAlaIleCysTrp 200
|||||
609 AGCCACCTGTCAAGACACCGGCTGCAAGGTTGGTCTAGCAATCTGTAT 658
201 AspMetArgPheProGlnLeuSerLeuLysLeuAlaGlnAlaGlyAlaGln 217
|||||
659 GACTGTGGGCTTCCTGAACTTCTTGAATTTGGCTCAAGCTGGGCGAGA 708
217 uLeuLeuThrTrpProSerAlaPheGlySerValThrGlyProAlaHis 234
|||||
709 AATPCTTACTTATTTCTCAGCCTTGGATCTGTACAGGCTCAGCCACT 758
234 rGlnValLeuLeuArgAlaArgAlaIleGlnSerGlnCysTrpValIle 250
|||||
759 GGGAGGTGCTGCTGGGCGCGCCCATTTGAATCTCAGTCTAATGATA 808
251 ALAAlaAlaGlnCysGlyArgHisGlnThrArgAlaSerTrpGlyHis 267
|||||
809 GCACACCGCAGTGTGAGCGCCACCATGAAACAGAGACATTTATGGCCA 858
267 sSerMetValValAspProTrpGlyThrValAlaAlaArgCysSerGln 284
|||||
859 TAGCATGTGTGTTACCCGTTGGGACACAGTGTGTGGCCGCTGCTCCAGG 908
284 LyrProGlyLeuGlyLeuAlaArgIleAspLeuHisPheLeuGlnGlnMet 300
|||||
909 GACCAGGCTCTGCTGCTGCAATGTATCTCACCATTCTTCTACACAGATG 958

301 ArgGlnHisLeuProValPheGlnHisArgProAspLeuTrpGlyLys 317
|||||
959 CGCCACACCTGCTGTTCTCCAGCACCGCAGACTGACTTATGAGCAG 1008
317 rLeuGlyHisProLeuSer 323
|||||
1009 TCTGGGTATCTCACCCTCT 1027
seq_name: gb-pr:AF069987
seq_documentation_block: 1385 bp mRNA linear PRI 23-JUL-1998
LOCUS AF069987
DEFINITION Homo sapiens nitrlase 1 (NIT1) mRNA, complete cds.
ACCESSION AF069987
VERSION AF069987.1 GI:3228665
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1385)
Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrilase and fhit homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
98337986
MEDLINE
2 (bases 1 to 1385)
Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
source
1. 1385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
1. 1385
/gene="NIT1"
77. 1060
/note="NIT1"
/codon_start=1
/product="nitrilase 1"
/protein_id="AAC39907.1"
/db_xref="GI:3228666"
/translation="MUGFTTRPHRFLSLCPGLRIPQLSVLCAPRRPRAAISSSSC
ELPLVAVCVSTSPDKQONFKTCAELVREARAGACIAPLPEAFDIARPAETLHLS
EPLGKLLLEVTOLAREGGLMTSLGFGHERGQWDEQTKYNGHVLLNSGAVATYRS
KTHLCVLEIYGQSGPMSNSTMPGSPLESFVSYPAGKIGLVADYDRPFLSLAQA
CAELLTYPSAFSGITPCRAHNEVILRRALLETQYVYVAAACGSHHRRKRSYSGHSMVD
PMGTIVARSGEGCLCLARDLNTYLQLRHRLHDFVQHRPDLTGNLGHPLS"
BASE COUNT 322 a 382 c 357 g 324 t
ORIGIN
alignment_scores:
Quality: 1487.00 Length: 327
Ratio: 4.828 Gaps: 2
Percent Similarity: 94.190 Percent Identity: 84.404
alignment_block:
US-09-357-675C-22 x AF069987 ..
Align seg 1/1 to: AF069987 from: 1 to: 1385
1 MetLeuGlyPheIleThrArgProHisGln.....LeuGly 14
|||||

77	ATCTGGGCTTCATCACAGGCTCTCTCAGATTCCGTCCCTTCGTGG	126
14	SThGlyTyrArgLeuLeuArgIleProValLeuCysThrGlnProArg	31
127	TCGTCGATCCGGATTACTCAACTCTGATCTTGTGTCAGCCACAGC	176
31	roaGlyIleMetSer...SerSerThrSerProGluLeuProLeuVal	46
177	CCAAAGCCATGGCTATCTCTCTCTCTCTGCAACTCCCTGGTGGCT	226
47	ValCysGlnValIleThrSerThrProAsnLysGlnIleAsnIlePhe	63
227	GTCGGCCAGTAACTATGACCCACAGACAGACAGACTTTAAACATG	276
63	salGluLeuValGlnGlnValAlaArgLeuGlyValAcCysLeuAlaPhe	80
277	TGCTGACTGTGTGAGAGCGTCGACAGCTGGTGTCCGCTGGCTTCC	326
80	eupProGluAlaPheAspPheIleLeuArgAsnProAlaGluThrLeu	96
327	TGCTGGAGGCAATTTGACTCTCTCACAGGAGCCCTGCAAGACGTAC	376
97	LeuSerGlnProLeuAsnGlyAspLeuLeuGlyIleIleThrSerGln	113
377	CTGTCTATACCACTGGGGGAAACTTTGGAAAGATACACCACTTCC	426
113	argGluCysGlyIleThrPheLeuSerLeuGlyIlePheIleGluArg	130
427	CAGGAATGTGTGACTCTGTGCTCTCTGTTGGTTCATGAGCTGGCC	476
130	IleAspThrGluGlnAsnGlnIleIleIleIleIleIleIleIleIle	146
477	AAGACTGGGACAACTCAGAAATCTAACATTTGACAGTCTGCTGAC	526
147	SerLysGlySerValValAlaSerTyrArgIleThrIleCysAspVal	163
527	AGCAAGGGGAGTAAAGTCCACTTACAGAAAGACACTCTGTGTGACT	576
163	IleGluIleProGlyGlnGlyProMetArgGluSerAsnTyrThrLys	180
577	AGAATTTCCAGGGGAGGGCTTATGTGTGAAGCAACTCTCAATGG	626
160	IleGlyThrIleLeuGluProProValLysThrProAlaGlyLysVal	196
627	GGCCCACTCTTGACTCACTGTGCAGACACACAGACAGCAATGGTCT	676
197	AlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysLeuAla	213
677	GCTGTCTGCTATGACATAGCGCTTCCCTGAATCTCTCTGGCATTTG	726
213	nalGlyValAlaGluIleLeuThrTyrProSerAlaPheGlySerVal	230
727	AGCTGGAGCAGAGATACTTACTTATCTTCAGCTTTTGATCCATTAC	776
230	LysProAlaIleHisThrGluValLeuLeuArgAlaAlaArgAlaIle	246
777	GCCAGGCCACTGGAGAGTGTGCTGGGGGCCCGGTGATGAAACCCAG	826
247	CysTyrValIleAlaAlaAlaGluCysGlyIleArgIleHisGluThr	263
827	TGCTATGTAAGGAGGACGACAGATGTGGAGGCCACCTGAAAGAGAGC	876
263	aseTyrGlyHisSerMetValValAspProThrGlyThrValValAla	280
877	AAGTTAAGGCCACAGCATGTGTGTGAACCTGGGGAAACAATGTGGCC	926
280	rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPhe	296
927	GCTGCTCTGAGGGCCAGGCTCTGCTTGGCCGAATAGACTCAACTAT	976
297	LeuGlnIleMetArgGlnIleHisLeuProValPheGlnIleAspArgPro	313
777	CTCGGCAAGTTGGCCGACACATCTCTGTGTTTCCAGCACCGCAGCGCTGA	1026

313 pleuArglySerLeuGlyHisProLeuSer 323
|||||
1027 CCTCTATGGCAATCTGGGTCCACCCACTGTCT 1057

seq_name: gb_ro:AF069985

seq documentation block:

LOCUS	AF069985	4481 bp	DNA	linear	ROD 23-JUL-1998
DEFINITION	Mus musculus nitrilase homolog 1 (Nlt1) gene, alternatively spliced product, complete cds.				
ACCESSION	AF069985				
VERSION	AF069985.1	GI:3242979			
KEYWORDS					
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 4481) Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tiliib,S., Dragescu,A., Wernuth,P., Rothman,J.H., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M. Nitrilase and fhl1 homologs are encoded as fusion proteins in <i>Drosophila melanogaster</i> and <i>Caenorhabditis elegans</i> Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)				
TITLE	2 (bases 1 to 4481) Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tiliib,S., Dragescu,A., Wernuth,P., Rothman,J., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M. Direct Submission Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA				
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
FEATURES					
Source	1..4481 Location/Qualifiers				
gene	/organism="Mus musculus" /db_xref="taxon:10090" /chromosome="1" /map="1q21-q23" join(541..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3944) /gene="Nlt1" /note="alternatively spliced" /product="nitrilase homolog 1" 541..3944 /gene="Nlt1" 541..606 /gene="Nlt1" /note="alternatively spliced" /number=1 join(605..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3658) /gene="Nlt1" /note="alternatively spliced" /codon_start=1 /product="nitrilase homolog 1" /protein_id="AAC40184.1" /db_xref="GI:3242980"				
exon	/translation="MLGFTTRPHDLCTGVRLLRIPIVLTCTPRPRRTSSSTSWELPLVAQCVISTPKNQENFKTCALVQEAARAGACLADEPAFDIARPAETLLSEPLNGLDLOYSQALARREGIWLISLGFHNRGQDWEONORIYNCHVLLNSGSVAASRKTHLGDVEIPGCGPMRESNYTKPGGTLPEPVTPYTPACKVGLATCYDMRPFELSLKLAQAGEILTVPSAFSGTSPAHMEVLLRLARAIIESOCYVIAAOCGRHETRASVYGSNVDPMGTVVANCSGEGGLARIDILHFLDQMQRHLPVQHRRRDLYLSGHHLS"				
exon	863..986 /gene="Nlt1" /note="alternatively spliced" /number=2 1177..1263 /gene="Nlt1" /number=3 1638..1889 /gene="Nlt1"				

exon	/number=4 2015..2118 /gene="Nt1"
exon	/number=5 2362..2495 /gene="Nt1"
exon	/number=6 2626..2751 /gene="Nt1"
exon	/number=7 3392..3944 /gene="Nt1"
BASE COUNT	1101 a 1132 c 1133 g 1115 t
ORIGIN	/number=8
alignment_scores:	Quality: 1347.00 Length: 828
	Ratio: 4.196 Gaps: 9
Percent Similarity:	38.768 Percent Identity: 38.406
alignment_block:	
US-09-357-675C-22 x AF069985	..
Align seg 1/1 to: AF069985 from: 1 to: 4481	
2 leuGlyPheIlethrArpProHisGlnLeuLeuCysThrGlyTyrar 18	
1178 CTGGGCTTCATCCACGAGCCTCTCTACCAACTCTGTGTACCGGATACGG 1227	
18 gLeuLeuArGliePrioValLeuCysThrGlnPro..... 29	
1228 ATTACTTCGAATCCAGTACTGTGTACTCAGCCAGGTAACATATTCGT 1277	
29 29	
1278 AATGTTCTCTCTGGCGCATACCTGCTTAGTTCGTGTATTAACTTCA 1327	
29 29	
1328 TAGTGGAGACAAAGATATCTCCAAACATGTAAGTGAACCTCCAGTAG 1377	
29 29	
1378 CTCTAGGCTTATTATGCTCTGTAAAGATGAACAACAGTGTGGGGTCA 1427	
29 29	
1428 CTGTAAATGCCAGACCTCGAGAAAAAGAGATCAACGAGATTAAAGT 1477	
29 29	
1478 CATCCACAGTTACATGAAAAGCCTGTCTGGGCTACAAAGACCGTCT 1527	
29 29	
1528 CAAAAAAGAAAAAGTCAAAAAGAAATTAAGTTCTGGCTCCACCCGAC 1577	
29 29	
1578 TTGAGCCTCTGAGTTTCTCTAGATAGAAAATCTGAGATACCTGCTGTG 1627	
30 Arg..ProArgThrMetSerSerSerThrSerTpgLLeuPro 43	
1628 TCATTACAGAGCCACAGACCATGTCTCTCATCACTCTCTGGAGCTGCC 1677	
44 leuValAlaValCysGlnValThrSerThrProAsnLysGlnLysAsnph 60	
1678 CTGGTGGCTGTGTGCCAGGTAAACATCAACACCAAGCAAGCAAGCAACTT 1727	
60 elysThrCysAlaGlnLeuValGlnGlnAlaAlaArgLeuGlyLacysL 77	

[illegible]

```

211 euAlaGlnAlaGlyAlaGlnIleuThrTyProSerAlaPheGlySer 227
|||||
2678 TGCGTCAAGCTGGGGGAGAGAAATACTTACTTATCCCTTACGCTTTGATCT 2727
228 ValThrGlyProAlaHisTrp..... 234
|||||
2728 GTTACAGGTCCAGCCCACTGGAGAGTAAGAGATGCCCTTCATTAACATA 2777
234 ..... 234
2776 TAGGCTCCCTTTAACTTACTTTAGTAGTGAGAGACAGTGTATCAAC 2827
234 ..... 234
2828 CTAATCCAGACACTTACGTGATAGCCACAGGTTCAAGCAGCATGCC 2877
234 ..... 234
2878 TTGACTAAAAAAGAAACAAATAGAACCTGGGCACTGGGGTGCAACC 2927
234 ..... 234
2928 TTTAATCCACACTCAGAGGACAGACAGAGTGATCTGTGATTGAG 2977
234 ..... 234
2978 GCCAGCTGCTCTACAGAGTGGGTTCCAAAGACAGCGGCTACACAGAG 3027
234 ..... 234
3028 AAACCTGTCTCAAAACAAAAACAAACCAACCACTTCAAAAGAGTGGG 3077
234 ..... 234
3078 CATGTGTGATGCTCTTTCATCTCAGCACTGTAGGCAAGGAGCAT 3127
234 ..... 234
3128 ATCTCTGGGTCTAGGCCAGCAGGTTACACAGGAGACCTTATCTCA 3177
234 ..... 234
3178 AAAGTACAAAACACAAACACTCTCTCTTCTTCTGATTTCTTCCTTTT 3227
234 ..... 234
3228 CATCTGGTGAAGAAATTCATTCTTACCTAGGTGAGTCACTGTTAAGA 3277
234 ..... 234
3278 GTTAGTTATTAACCTAGTTGTTAGTGAGAAAGATGGGCCACAGTGC 3327
234 ..... 234
3328 CGCTTCTTGCACTTGCTATTAACGCTGATGTCTTGAGATCAACGCA 3377
235 ..... 235
3378 CGTCTACTCTCAGGTGCTGCTGGGGCCGCGCATTTGAATCTCAGTGC 3427
248 TyrValIleAlaAlaAlaGlnCysGlyArgHisHisGlnThrArgAlaSe 264
|||||
3428 TATGTATATAGACAGCAGGAGTGTGAGCCACCATGAAACAAGAGCAG 3477
264 rTyGlyHisSerMetValValAspProTrpGlyThrValAlaAlaArgC 281
|||||
3478 TTTAGGCCATATGATGTGTGTGACCCCTGGGCAAGTGTGGCCGCT 3527
261 ysserGlnGlyProGlyLeuGlyLeuAlaArgIleAspLeuHisPheLeu 297
|||||
3528 GCTCGAGGAGGAGGAGGCTTGCCTTGCCTGCAATTAATCTCACTTCTCA 3577

```

```

298 GlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAspLe 314
|||||
3578 CAACAGATGCCGCAACACCTGCTGTCTCCAGCAGCCGACGCTGACCT 3627
314 uTyGlySerLeuGlnHisProLeuSer 323
|||||
3628 CTATGGCAGTCTGGGTCACTCCACTCTCT 3655
seq_name: gb_hlg:AC084821
seq_documentation block:
LOCUS AC084821 211771 bp DNA linear HTG 29-JAN-2002
DEFINITION Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE,
1 ordered pieces.
ACCESSION AC084821
VERSION AC084821.24 GI:18390260
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211771)
REFERENCE 1 (bases 1 to 211771)
AUTHORS Deschamps, S., Gu, W., and Roe, B. A.
JOURNAL Mus musculus BAC Clone rp23-395h6
REFERENCE 2 (bases 1 to 211771)
AUTHORS Deschamps, S., Gu, W., and Roe, B. A.
JOURNAL Unpublished
TITLE Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jan 29, 2002 this sequence version replaced gi:16076995.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 211771: contig of 211771 bp in length.
* Location/Qualifiers
1. 211771
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone_lib="rp23-395h6"
/clone_lib="RP2 Mouse BAC Library 23"
BASE COUNT 55622 a 48688 c 49207 g 56231 t 23 others
ORIGIN
alignment_scores:
Quality: 1343.00 Length: 824
Ratio: 4.197 Gaps: 9
Percent Similarity: 38.835 Percent Identity: 38.471
alignment block:
US-09-357-675C-22 x AC084821 ..
Align seg 1/1 to: AC084821 from: 1 to: 211771
2 LeuGlyPheIleThrArgProProHisGlnLeuGlyThrGlyTyArg 18
|||||
529 CTGGGCTTACACACGCTCTCCACCACTCTGTGTACCGGATACCG 578
18 glauLeuArgIleProValLeuGlyThrGlnPro..... 29

```

```

|||||  |||||||  |||||||
579 ATACTTCGAACCCAGTACTTTGTACTCAGCCAGTAACATATTTCGT 628
29 ..... 29
629 AATGTCCTCTCTGCGGCATAGCTGCTAGTGTGTTATTATTAATTCA 678
29 ..... 29
679 TAGTGGAGACAAGATATCTCCAAACATGTATAGTGAATCTCAGTAG 728
29 ..... 29
729 CTCTAGGGTTATTATGCTCTGTAGAAATGAAACAAAGTGTCTGGGGTCA 778
29 ..... 29
779 CTGTAATGCCAGACATCGAAAAAAGAGATCAACAGAGATTAAAGT 828
29 ..... 29
829 CATCCCAAGTTACATAGAAAGCGCTGTCTGGGTACAAAGACCGTGTCT 878
29 ..... 29
879 CAAAAAGAAAAAGTCAAAAAAATRAAGTTTCTGCTCCACCCCCAC 928
29 ..... 29
929 TTGAGCCTTGAGTTTCTCTAGATAGAAATCTAGGATACCTGCTGTG 978
30 .....Arg..ProArgThrMetSerSerThrSerThrProAlaLeuPro 43
979 TCATTTCAGAGCCGAGAACCATGTCATCACTTCCTGGAGCTGCC 1028
44 LeuValAlaValCysGlnValThrSerThrProAlaLeuValAlaLeu 60
1029 CTGTGGCTGTGTGCTCAGGTACATCAACCAACAAAGCAAGAACTT 1078
60 eLysThrCysAlaGluLeuValGlnValAlaArgLeuGlyAlaCysL 77
1079 TAAACATGTCTGAGTGGTTCAAGAGGCTGCCAGACTGGTGTCTCC 1128
77 euAlaPheLeuProGluAlaPheAspPheIleAlaArgAsnProAla 93
1129 TGGCCTTTCGCTGAGGCAATTGACTTTATTGCACGAACCTGCCGAG 1178
94 ThrLeuLeuSerGluProLeuAsnGlyAspLeuGlnGlyIntYrSe 110
1179 ACATTACTCTCTGCTCGAACCACTGAATGGGATCTTTGGCCAAATATAG 1228
110 rGlnLeuAla..... 113
1229 CCAGCTTGCCAGGTACAAGGTAGGGTTTGGGGGGGGGGTGTGTGT 1278
113 ..... 113
1279 GTAGACTCCCTGTTGAGCAGTGTCTGATGCCACAGAGGGGGTGA 1328
114 .....ArgIuCysGlyIleTyrPle 120
1329 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1378
120 uSerLeuGlyGlyPheHisGluArgGlyGlnAspTyrPleGluGlnAsn 137
1379 GTCTTGGGCGGTTTCCAGAGCGTGGCCAAAGACTGGAGAGCAATCAGA 1428
137 ySileTyrAsnCysHisValLeuLeuAsnSerLysGly..... 149
1429 AAATCTACAATGTATGTGCTTTTGAACAGCAAGGGTGAGACGTATAG 1478
149 ..... 149
1479 CCCTTAGTTGCTCTTCCAGGCTTCTCTACCTTCTCTCAACTCAGATC 1528
149 ..... 149
1529 TTCAGCATTTCTTTCTCAACTGCTTGGCCTTGGGGTTTATGCTTGT 1578
149 ..... 149
1579 CTCAGCTTACAAGTTACTTTGGGAGAGAGCTAAGCTGACTTGTAGA 1628
149 ..... 149
1629 GTGTCTGACTGAGATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1678
150 .....Ser..ValValAlaSerTyrArgLysTh 158
1679 CTCTCTCTCTCTCTCTCTCTCTAGATCAGTACGTAGCGCCAGTACAGGAAAC 1728
158 rHisLeuCysAspValGluIleProGlyGlnGlyProMetArgGlySer 175
1729 ACATCTGTGCATGTAGAGATCCAGGTGAGGGGCGCATGAGAGAAAGCA 1778
175 snTyrThrLysProGlyGlyThrLeuGluProProValLysThrProAla 191
1779 ACTATACCAAGCCTGAGAGCACTCTTGAGCCACCTGATCAAGACACGGCT 1828
192 Gly..Lys..... 193
1829 GGCAAGGTGGAGTTTGGAAAGAAATGGACTATGTGAAAAAGGTATA 1878
193 ..... 193
1879 CACCCCATTTGGTGACGCTAATGAAGCACTGCTTCTGTGAATG 1928
194 .....ValGlyLeuAlaI 198
1929 GGACAGGAAGATGATGGGTGTTTCTCATTTTCAGTTGGTGTGCAAT 1978
198 eCysTyrAspMetArgPheProGluLeuSerLeuLysLeuAlaGln 215
1979 CTGTATACATGCGGTTCCCTGAACCTTTTGAATTTGGCTCAAGCTG 2028
215 LysAlaGluIleLeuThrTyrProSerAlaPheGlySerValThrLys 231
2029 GGGCAAAATACTTACTTATCTTCAGCTTGGATGATGTACAGGTCCA 2078
232 AlaHisTyr..... 234
2079 GCCCACTGGAGGTAAAGAAATGCTCTTTCATATATAGGCTTCCTT 2128
234 ..... 234
2129 AACCTTACTTTAGTGTGGAGCAGAGTGTACAAACCTAATCCAGC 2178
234 ..... 234
2179 ACTTCAGTGATAGCCACAGTTCAAGCCAGCATGCTTGACTATAAAAA 2228
234 ..... 234
2229 GAAACAAATATGAAGCTGGGCAGTGGGGGTGCACACCTTATATCCAGC 2278
234 ..... 234
2279 ACTCAGAGAGCAGAGCAGGTGATCTGAGTTTGAGGCCAGCGTGTG 2328
234 ..... 234
2329 TACAGATGGGTTCCAGAGACAGCGGTACACAGAGAAACCTGTCTC 2378
234 ..... 234
2379 AAAAAAAGCAACCCACACCTTCAAAAGTGGGCAATGTGTGTGCA 2428

```

```

234 ..... 234
2429 TGCTCTCATCTCAGACACTGTGAGCAAGCAAGGACAGATCTGTGGGT 2478
234 ..... 234
2475 CTAGACGACGAGGGTTACACAGGAGACCTTATCTCAAAAAGTCAAAAC 2528
234 ..... 234
2525 ACAAACTCCTCTTCTTCTACTGATTTCTTCTTCTTCTTCTGAGG 2578
234 ..... 234
2579 AAATTCATTTCTTAGTAGTAGTGATCATGTTAAGAGTTAGTTATTA 2628
234 ..... 234
2629 CTCTAGTTGGTTAGTAGAAGATTGGCCACAGTCCCGCTTCTGTCA 2678
235 ..... G1 235
2679 CTTCCTATTAAACGCTGTATGTTCTGAGTCAACGACGCTCATCTC 2728
235 u.valleuleuAryAlaAryAlaIleGluSerGlnCysTyryValIleAla 251
2729 AGGTGCTGTGCGGGCCGCCCATGTGATCTGATGTAATGTAATGCA 2778
252 AlAlaGlnCysGlyAryGlnHisGluThrAryAlaSerTyryGlnHis 268
2779 GCACGCGAGTGTGACGCCACCATGAACAAAGACAAAGTATGCGCATAG 2828
268 rmetValValAspProTrpGlyThrValAlaAryCysSerGlnGlyP 285
2829 CATGTGTGTGACCCGTTGGGCGACAGTGTGCCCCGCTGCCGAGGAC 2878
285 roGlyLeuCysLeuAlaAryIleAspLeuHisPheLeuGlnMetAry 301
2879 CAGGCTCTGCTGCTGCGAATTGATCTCCACTTTCACAAACATGCGC 2928
302 GlnHisLeuProValPheGlnHisAryAryProAspLeuTyryGlySer 318
2929 CAACACCTGCTGTGTTCAGACCGCAGACCTGACCTATGACAGTCT 2978
318 uGlyHisProLeuSer 323
2979 GGGTCATCCACTCTCT 2994

```

seq_name: gb_htg:AC087229

```

seq_documentation_block:
LOCUS AC087229 213203 bp DNA linear HTG 05-FEB-2002
DEFINITION Mus musculus chromosome 1 clone rp23-191a19, WORKING DRAFT
ACCESSION AC087229
VERSION AC087229.15 GI:18497119
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Jiang X., Song L., Gu W. and Roe B.A.
TITLE Unpublished
JOURNAL 2 (bases 1 to 213203)
REFERENCE
AUTHORS Jiang X., Song L., Gu W. and Roe B.A.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Feb 5, 2002 this sequence version replaced g1:18390237.

```

```

----- Genome Center
Center: Department of Chemistry And Biochemistry
The University of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11373: contig of 11373 bp in length
* 11374 11473: gap of unknown length
* 11474 43477: contig of 32004 bp in length
* 43478 43577: gap of unknown length
* 43578 126526: contig of 82949 bp in length
* 126527 213203: gap of unknown length
* 126627 213203: contig of 86577 bp in length.
*
Location/Qualifiers
source 1. 213203
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-191a19"
/clone_lib="RPI Mouse BAC Library 23"
BASE COUNT 57858 a 50465 c 49150 g 55425 t 305 others
ORIGIN
alignment_scores:
Quality: 1320.50 Length: 824
Ratio: 4.127 Gaps: 10
Percent Similarity: 38.835 Percent Identity: 38.350
alignment_block:
US-09-357-675C-22 x AC087229 ..
Align seg 1/1 to: AC087229 from: 1 to: 213203
2 LeuGlyPheIleThrAryProTrpHisGlnLeuLeuCysThrGlyTyr 18
48196 CTGGGCTTCATACACACGCTCCACCAACTCTCTGTGTACCGATACCG 48245
18 gLeuLeuAryIleProValLeuCysThrGlnPro..... 29
48246 ATTACTTCGAACCCAGTACTTGTACTCAGCCAGGTAACATATTTCGT 48295
29 ..... 29
48296 AATGTTCTCTCTGCGGCATAGCTGTTAGTTCGTTATTAACTTCA 48345
29 ..... 29
48346 TAGTGGAGACAAGATATCTCCAAACATGTAAAGTGAATCCAGTAG 48395
29 ..... 29
48396 CTCTAGGGTTATTATGCTCTGTGAAGATGAACAACTTCTGGGGTCCAC 48445
29 ..... 29
48446 CTGTAAATGCAGACTCGAAAAAAGAGAGATCAACCAAGAGTTTAAGT 48495
29 ..... 29
48496 CATCCCGATTACATAGAAAGGCTGTGTGGCTACAAAAAGACGTCCT 48545
29 ..... 29
48546 CAAAAAAGAAAGTCAAAAAAATAAAGTTTGTGGCTCCACCCCCAC 48595

```

```

29 ..... 29
48396 TTGAGCCTCTGATTTCTCTAGATAGAAATCTGAGATACCTGCTGTG 48645
30 .....Arg..ProArgThrMetSerSerSerThrSerTrpGluLeuPro 43
48646 TCATTCAAGAGGCCAGAACCATGCTCTCATCACTCTCTGGAGAGCTGCC 48695
44 LeuValAlaValCysGlnValThrSerThrProAsnLysGlnGluAsnPh 60
48696 CTGGCGCTGTGTGCCAGGTAACTCAACACCAACCAAGCAAGAACACTT 48745
60 elysThrCysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysL 77
48746 TAAACATGCTGAGTGTGTTCAAGAGGCTGCCAGACTGGGTGCTTCC 48795
77 euAlaPheLeuProGluAlaPheAspPheIleAlaArgAsnProAlaGlu 93
48796 TGGCCTTCTGCTGAGGCATTGTGACTTTATTGCGAACCCTGCCAG 48845
94 ThrLeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyLysTyrSe 110
48846 ACATTACTCTGCTCCGAACCACTGATGGGATCTTTGGGCCAATATAG 48895
110 rGlLeuAlaArg..Glu..... 115
48896 CCAGCTTCCAGGTACAAAGGTAGGGTTGGGGGGGGGTTGTGTG 48945
115 ..... 115
48946 TAGAACCTCCGTGTTGAGCAGTGTCTGATTTGCCACAGAGGGGTTAG 48995
116 .....CysGlyLLeu 120
48996 TTGTGTGTGTGTGTCCTCCGCCGCCGCCGCCAGGGAATGTGAATCTGG 49045
121 SerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnAsnGlyL 137
49046 TCTTGGGCGGTTTCCACGAGCGTGGCCAAAGACTGGGAGCAAAATCGAA 49095
137 sIlLeuTyrAsnCysHisValLeuLeuAsnSerLysGly..... 149
49096 AATCTACAAATTGTGATGTGCTTTTGAACAGCAGGGTGAAGCTATAGC 49145
149 ..... 149
49146 CCTAGTTTGCCTTCCAGGCTTCTACTTTCCTCAACTCAGATCT 49195
149 ..... 149
49196 TCAGCATTTCTTCTCAACTGCTTGCCCTTGGGGTTTATGCTCTCTTC 49245
149 ..... 149
49246 TCAGCCTACAGTTACTCTTGGGAGAGAGCTAGAGCTGAGCTGTAGAG 49295
149 ..... 149
49296 TGTCTGTACTAGTATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 49345
150 .....Ser..ValValAlaSerTyrArgLysThr 158
49346 TCTCTCTCTCTCTCTCTAGATAGTAGTGCCAGTTACAGAGAGACA 49395
159 HisLeuCysAspValAluIleProGlyGlnGlyPrometLarGluSerAs 175
49396 CATGTGTCGATGTAGAGATCCAGGTACGAGGCGCGATGAGAGAAAGCA 49445
175 nTyThrLysProGlyGlyThrLeuGluProProValLysThrProAlaG 192
49446 CTATACCAAGGCTGGAGGACACTCTTGAGCCACTGTCAAGACACCGGTG 49495
192 Ty.Lys..... 193

```

```

49496 GCAAGGTGGAGATTGGAAAGAAATGGGACTATGTGAAAGTGTATAC 49545
193 ..... 193
49546 ACCCCCATTTGGTGCAGGCTAATGAAGCACTAGTGTCTGTGAATGG 49595
194 .....ValGlyLeuAlaIle 198
49596 GACAGGGAAGATGAGTGGGTGTTCTCATTTTCAGGTGTGCTTACCAATC 49645
199 CysTyrAspMetLarPheProGluLeuSerLeuLysLeuAlaGluAlaG 215
49646 TGTATGACATGCGGTTCCCTGAACCTTCTTTGAATTTGGCTCAAGCTGG 49695
215 yAlaGluIleLeuThrTyrProSerLarPheGlySerValThrGlyPro 232
49696 GGCAGAAATACTTACTTATCCTTCAGCCCTTGTGATCTGTACAGGTCCAG 49745
232 IaHisTrp..... 234
49746 CCCACTGGAGGTAAAGAAATGCTCTTCATATATATAGGCTCTCTTA 49795
234 ..... 234
49796 ACCTTACTTTTAGTGTGGAGACAGTGTGTACAAACCTAATCCAGACA 49845
234 ..... 234
49846 CTTAGTGGATAGCCACAGGTTCAAAAGCAGCATGCTTGACTAAAAAG 49895
234 ..... 234
49896 AAAACAAATAGAACCTGGGAGTGGGGGTGCACACTTTAATCCAGACA 49945
234 ..... 234
49946 CTCAGAGAGCAACACAGGTGATCTGTAGTTTAGGCCACCTGTGTCT 49995
234 ..... 234
49996 ACAGAGTGGGTTCCAGAGACGCCAGGGCTACACAGAGAAACCTGTCTCA 50045
234 ..... 234
50046 AAAACAAAAGCAACCAACACCTTCAAGAGTGGGCAATGGTGTGCAT 50095
234 ..... 234
50096 GCTTTCATCTCAGACATGTGAGCAAGGACGACGATCTGTGGGTTC 50145
234 ..... 234
50146 TAAGCCAGCAGGTTACACAGAGGAGACTTATCTCAAAAGTAAACA 50195
234 ..... 234
50196 CAACACCTCCCTTCTTACTGATTTCTTCTTCTTCTGAGAGA 50245
234 ..... 234
50246 AATTATTTCTTAGCTAGGTGTGAGTCATGTTAAAGATTAATTAAC 50295
234 ..... 234
50296 TCTAGTGTAGTGAAGATTGGGCCACAGTGCCGCTCTGTGTAC 50345
235 .....Glu 235
50346 TTGCCTATTAAACGCTGATGTCTGTGAGTCAACGACGCTCATCTCA 50395
236 ValLeuLeuArgAlaIleGluSerGlnCysTyrValIleAla 252

```

```

50396 GGTCTGCTGCGGGCCGCCCATTTGATCTCAGTCTATGTAATAGCAG 50445
252 1aaIaagInCySgLYa9HishIsLwHrAaGAlaseTyGlyHisSer 268
|||||
50446 CAGCGCAGTGTGGACGCACCATGAAACAGACAGCAAGTATGGCATAGC 50495
269 MetValValAspProTTPGlyThrValValAlaArgCysSerGluGlyPr 285
|||||
50496 ATGTGGTGTGACCCGCGGCGACAGTGTGGCCCGCTGCTCCGAGGAGCC 50545
285 oGlyLeuCySneulAlaArgIleAspLeuHishPheLeuGlnImeArGg 302
|||||
50546 AGCGCTCTGCTTCGAAATTCATCTCCACTTCTACACAGATGCGCC 50595
302 InhIsLeuProValPheGlnHisArgArProAspLeuTyGlySerLeu 318
|||||
50596 AACACCTGCGCTGTTCACAGACCGCAGACTGACTTATGACAGTCTG 50645
319 GlyHisProLeuSer 323
|||||
50646 GGTATCTCCTCTCT 50660

seq_name: gb_p1:AF069984
seq_documentation_block: 4079 bp DNA linear PRI 23-JUL-1998
LOCUS AF069984
DEFINITION Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced
product, complete cds.
ACCESSION AF069984
VERSION AF069984.1 GI:3242977
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4079)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrilase and fhl homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
JOURNAL
MEDLINE
98337986
2 (bases 1 to 4079)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Yench St., Philadelphia, PA 19107, USA
FEATURES
source
1..4079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3267)
/gene="NIT1"
/note="alternatively spliced"
/product="nitrilase homolog 1"
635..3569
/gene="NIT1"
/note="NIT1"
/gene="NIT1"
/note="alternatively spliced"
/number=1
join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3267)
/gene="NIT1"
/note="NIT1; alternatively spliced"
/codon_start=1
/product="nitrilase homolog 1"

```

```

/protein_id="AAC39901.1"
/db_xref="GI:3242978"
/translation="MIGFTTRPPHRLSLICPGLRIPOLSVLCAQPRPRAMAISSSC
EPLVAVCOVSTSPKQONFKTCAELVREARADLAPLPAEPFIADPAETLHL
EPLGKILEEYRDLAREGIMLSIGFHEHGGDMROTKIVNCHVLNKGAVATYR
KTHLCEVETPGGPRMCESENSTMPGSLSEPVSTPAKIGLAVCYMKRPPELSIALQA
GAEIITPSPAFSGITGPAAHEVILRRARILETCYVAAACGRHHEKRASYSCHSNVD
PMTGVVARSSEGPGLARIDLNLRLRLPVRQRRRDLXGNLGHPLS"
681..711
/gene="NIT1"
/note="1A: alternatively spliced"
712..741
/gene="NIT1"
/note="1B: alternatively spliced"
940..1053
/gene="NIT1"
/note="1C: alternatively spliced"
1288..1383
/gene="NIT1"
/number=2
1636..1890
/gene="NIT1"
/number=3
2011..2114
/gene="NIT1"
/number=4
2321..2454
/gene="NIT1"
/number=5
2609..2734
/gene="NIT1"
/number=6
3001..3569
/gene="NIT1"
/number=7

BASE COUNT 940 a 1082 c 1043 g 1012 t 2 others
ORIGIN

alignment_scores:
Quality: 1211 50 Length: 660
Ratio: 3.946 Gaps: 10
Percent Similarity: 46.515 Percent Identity: 41.667

alignment_block:
US-09-357-675C-22 x AF069984 ..
Align seg 1/1 to: AF069984 from: 1 to: 4079
2 LeuGlyPheIleThrArgProProHisGln.....LeuLeuCyStH 15
|||||
1289 CTGGGCTTCATCACCAGGCGCTCTCACAGATTCCTGTCCTGTGTC 1338
15 rGlyTyArgLeuLeuArgIleProValLeuCyStHGlPro..... 29
|||||
1339 TGGAGTCGCGGATACCTCACTCACTGATCTTGTGTCAGCCAGGTAAC 1388
29 ..... 29
1389 ACGTTTGTGTGTCTCAGTGCCTGGCAGCTTAGATGCTAGTTGTTAA 1438
29 ..... 29
1439 ATGATAGTGGAGACACAGAGAGTGTCAACTATCCACACATTGATTGGT 1488
29 ..... 29
1489 GAGCCCTACTAGCCCTGGTCAACGTGCCCTGTAAAGCATGATCAAAAG 1538
29 ..... 29
1539 GAAGTCAGCTTTCCTGCTCCACTTGGACCCCTTAGCATTAATTG 1588

```



```
105458 TAGCAATTACCAAAATAGTACAATGGTAGATGCTGTAATGCC 105507
237 ..... 237
105508 TCACCTGTCACCTCCCACTATTGCTACATGTAAGTAACACACA 105557
238 ..... LeuArgAlaArgAlaIleGluSerGlnCys 248
105558 TCTCATGCCAGGTGTGCTGCGGGCCCTGCTATGAAACCCAGTGTCTA 105607
248 rValIleAlaAlaIleGlnCysGlyArgHisHisGluThrArgAlaSer 265
105608 TGTAGTGCAGCAGCAGCAGTGTGAGCCACCATGAGAGAGACAAGT 105657
265 yrcGlyHisSerMetValValAspProTyrGlyThrValValAlaAlaG 281
105658 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGGCCCTGC 105707
282 SerGluGlyProGlyLeuGlyLeuAlaArgIleAspLeuHisPheLeuG 298
105708 TCTGAGGGGCGCAGGCTCTGCTTGCCTTGCCTGCAATAGACCTCACT 105757
298 nGlnMetArgGlnHisIleuProValPheGlnHisArgArgProAspLeu 315
105758 ACAGTTGCCCGCACACCTGCTGTGTTCCAGCACCGAGGCTGACCTCT 105807
315 yrcGlySerIleuGlyHisProLeuSer 323
105808 ATGGCAATCTGGGTCAACCCACCTGCT 105833
seq_name: gb_hlg:AL354714_5
seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome 1)
alignment_scores:
Quality: 1211.50 Length: 660
Ratio: 3.946 Gaps: 10
Percent Similarity: 46.515 Percent Identity: 41.667
alignment_block:
US-09-357-675c-22 x AL354714_5 ..
Align seg 1/1 to: AL354714_5 from: 1 to: 110000
2 LeuGlyPheIleThrArgProHisGln.....LeuLeuGlySer 15
|||||.....|
3858 CTGGGCTTCATCACCGAGGCTCTCACAGATTCCTGCTCTGTGTCTC 3907
15 rGlyTyrArgLeuLeuArgIleProValLeuGlySerHisPro..... 29
|||||.....|
3908 TGGACTCCGATACCTCACTCACTCACTTGTGCTGACGCCAGGTAAAC 3957
29 ..... 29
3958 ACCTTTTGTGTTCTCCTCAGTGCCTGCACTTATGATGCTCAGTTGTAA 4007
29 ..... 29
4008 ATGATATAGTGGAGACACAGAGGTGTCAATATCCACACATTTGATTGT 4057
29 ..... 29
4058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGACATGATCAAAAG 4107
29 ..... 29
4108 GAATGCCAGCTTTCCTGCTCTCCACTTGCACCCCTTAGATTAATTTG 4157
30 ..... ArgPr 31
4158 CTTCCTGCTGCTATGAATCTGAGAACTGCTGCTATGCTGTTCACAGCC 4207
31 oArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 47
|||||.....|
4208 CAGAGCCATGCTATCTCTCTCTCTCTCTGGAACCTGCCCTGGTGGCTG 4257
47 aLcysGlnValIleThrSerThrProAsnLysGlnIleAsnPheLysThr 63
|||||.....|
4258 TGTGCCAGTAAACATCGACCCAGACACAAGCAAGCAACTTAAACATGT 4307
64 AlAGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
|||||.....|
4308 GCTGAGCTGTGTGAGAGGCTGCCAGACTGGTCTGCTGCTGCTGCTTCT 4357
80 uProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 97
|||||.....|
4358 GCCTGAGGCACTTTCCTTCACTTGTGACGGAGACCTGCAAGACGCTACAC 4407
97 euSerGluProLeuAsnGlyAspLeuLeuGlyGlnIleYrSerGlnLeuAla 113
|||||.....|
4408 TGTCTGAACCACTGGTGGGAACTTTTGAAGAAATACACCACTTGCC 4457
113 ..... 113
4458 AGGTATCAGGGAATAGCGAGGAGAGTAGAATCTTGTGTGACAGTGT 4507
113 ..... 113
4508 CCCTGGCTCCAGATATGAGAGGTAGAGCCCTTGGAAGTCACTAGTAAG 4557
114 ..... ArgGluCysGlyIleTrpLeuSerLeuGly 123
4558 TGTACAGTGTCTTCCCTCCCTCCAGGAATGTGACCTGCTGCTTGGGT 4607
124 GlyPheHisGluArgGlyGlnAspTrpGluGlnAsnGlnIleIleTyrAs 140
|||||.....|
4608 GGTTCATGAGCGGTGGCCAAAGACTGGAGAGACTGAGAAATCTACAA 4657
140 nCysHisValLeuLeuAsnSerLysGly..... 149
4658 TTGTACGCTGCTGCTGACACGAAAGGTGAGACTTTTAACCTTTAGCC 4707
149 ..... 149
4708 TGCCTTCCCATGCTCTTCTACCTAGATTCTCCAGATTGTTCTCAAC 4757
149 ..... 149
4758 TCTTATTCCTTGACCAAGGATTTAGGGGTGTCTACTTCACTTCTCTTA 4807
149 ..... 149
4808 GCCTATTAACATCTCTCTCTGAGAGAGATGACCAAGCTTCTAGACA 4857
150 ..... Ser...ValValAlaSer 155
4858 CCAAGCATGATATTCTTCTTCTTACTGTAAGGGGAGTGTGTCGCACT 4907
155 yArgGlyThrHisLeuGlyAspValGluIleProGlyGlnGlyPromet 171
|||||.....|
4908 ACAAGAGACACATCTGTGACGTAGATTCACAGGCGAGGCGCTATG 4957
172 ArgGluSerAsnTyrThrLysProGlyGlyThrLeuGluProProValLy 188
|||||.....|
```

```
4958 TGTGAAGCAACTCTACCATCCTGGGCCAGTCTTGATCATCTGTAG 5007
188 sThpProiaGlyValGly Leu ..... 196
5008 CACACCGACGACAGGATGAGATTGTGAAGGATGAGGAGGACAG 5057
196 ..... 196
5058 GAATACCTTGAAGTGGCAGTAGAGATGAAAGCCCTAAGAGAGGGGTA 5107
196 ..... 196
5108 ATGGAATATGACTAGATGCTGTGCAAAACAGAGCAGGAGACTACTAG 5157
197 ..... AlaIcysTyrAspMetArg 203
5158 TAGCTGTTTTCATTCAGATTGGTCTAGCTGTCTGTATGACATGGGG 5207
204 PheProGluSerLeuLeuAlaGlnAlaGlyAlaGluLeuLeuTh 220
5208 TTCCTGAACTCTCTGTCGATGGCTCAAGCTGGAGCAGAGATCTTAC 5257
220 rTyProSerAlaPheGlySerValThrGlyProAlaHisTrpGluVal 236
5258 CTATCCTTACGCTTTGGATCATTCATACAGGCCACCCACTGGGAGTTAA 5307
237 Leu ..... 237
5308 GATGATGCTTTTAAACATAAGGCTTTTCTTAACCTACTTCCCC 5357
237 ..... 237
5358 CCTTGGCCCTACAGTTAAATTCCTCCCTTCCACCTAATGGGAAAC 5407
237 ..... 237
5408 TCATTCGCCAGATATTTCTCTCATGAAATAGTAAATAGTAATCATTC 5457
237 ..... 237
5458 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGCTGTAATGTCCC 5507
237 ..... 237
5508 TCACCTGTCACTCCCACTATTGGCTACATGTACTTAAGTAACACACA 5557
238 ..... LeuArgAlaArgAlaIleGluSerGlnCysTy 248
5558 TCTCATGCCCGAGTGTGCTGCGGGCCCTCTATCGAAACCACTGCTA 5607
248 rValIleAlaIleAlaGlnCysGlyArgHisIleGluThrArgAlaSerT 265
5608 TGTAGTGGCAGACACAGTGTGACGCGCCACCATGAGAAAGACCAACTT 5657
265 rGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 281
5658 ATGGCCACAGCATGGTGTAGACCCCTGGGAAAGTGTGGCCCGCTGC 5707
282 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeuG 298
5708 TCTAGAGGGGCGAGGCTCTGCTTGCCTGCGGAATAGACTCACTATCTG 5757
298 nGluMetArgGlnHisLeuProValPheGlnHisArgArgProAspLeuT 315
5758 ACAATTGGCGGACACCTGCTGTGTCTCAGACACCGCAGGCGCTGACTCT 5807
315 rGlySerLeuGlyHisProLeuSer 323
5808 ATGGCAATCTGGGTACCCACTGTCT 5833
seq_name: gb_htg:AL590651
seq_documentation_block:
```

```
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
ACCESSION AL590651
VERSION AL590651.4 GI:13990236
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Harrison,E.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CH10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequests@sanger.ac.uk
On May 7, 2001 this sequence version replaced gi:13751001.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA137A12
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 164729 bases at least Q40
Consensus quality: 165486 bases at least Q30
Consensus quality: 166169 bases at least Q20
Insert size: 167063; sum-of-contigs
Insert size: 171878; 3.4% error; agarose-fp
Quality coverage: 8.28x in Q20 bases; sum-of-contigs quality
coverage: 8.11x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 9373: contig of 9373 bp in length
* 9374 9473: gap of 100 bp
* 9474 29891: contig of 20418 bp in length
* 29892 29991: gap of 100 bp
* 29992 36239: contig of 6248 bp in length
* 36240 36339: gap of 100 bp
* 36340 103784: contig of 67445 bp in length
* 103785 103884: gap of 100 bp
* 103885 114878: contig of 10994 bp in length
* 114879 114978: gap of 100 bp
* 114979 126347: contig of 11369 bp in length
* 126348 126447: gap of 100 bp
* 126448 137764: contig of 11317 bp in length
* 137765 137864: gap of 100 bp
* 137865 163160: contig of 25296 bp in length
* 163161 163260: gap of 100 bp
* 163261 167863: contig of 4603 bp in length.
Location/Qualifiers
1. 167863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-137A12"
/clone_1bp="RPC1-11.1"
1. 9373
/note="assembly_fragment:02360
fragment_chain:1
clone_end:T7
vector_side:left"
```

```
misc_feature      9474..29891
/note="assembly_fragment:01854
fragment_chain:1"
misc_feature      29932..36239
/note="assembly_fragment:00805
fragment_chain:1"
misc_feature      36340..103784
/note="assembly_fragment:01467
fragment_chain:1"
misc_feature      103885..114878
/note="assembly_fragment:00991
fragment_chain:2"
misc_feature      114979..126347
/note="assembly_fragment:01503
fragment_chain:2"
misc_feature      126448..137764
/note="assembly_fragment:01297
fragment_chain:2"
misc_feature      137865..163160
/note="assembly_fragment:00117
fragment_chain:2"
misc_feature      163261..167863
/note="assembly_fragment:01267
fragment_chain:2"
clone_end:SP6
vector_side:right"
BASE COUNT      43927 a 39935 c 40002 g 43197 t      802 others
ORIGIN

alignment_scores:
    Quality: 1211.50      Length: 660
    Ratio: 3.946      Gaps: 10
    Percent Similarity: 46.515      Percent Identity: 41.667

alignment_block:
US-09-357-675c-22 x AL590651/rev ..

Align seg 1/1 to reverse of: AL590651 from: 1 to: 167863

      2 LeuGlyPheIleThraProProHisGln.....LeuLeuCysThr 15
      |||||
151392 CTGGGCTTCATCACCAGGCTCCTCTCAGATTCCTGCTTCCTGTC 151343
      15 rGlyYrAglLeuArlleProValLeuCysThrGlnPro..... 29
      |||
151342 TGGACTCCGGAATACCTCAACTCTCACTTGTGCTCAGCCAGGTAC 151293
      29 ..... 29
151292 ACCTTTTGTGTGTCCTCAGTGCCTGGCCTAGATGCTCAGTTGTTAA 151243
      29 ..... 29
151242 ATGGATAGTGGAGACACAGAGTGTCACTATCAACACATTGATTGCT 151193
      29 ..... 29
151192 GAGCCCTACTAGCCCTGGGTCAACGTGCTGTAGAGCATGATCAAAAG 151143
      29 ..... 29
151142 GAAGTCCAGCTTTCCTGCTCTCCACTTGACCCCTTAGCATTAATTGG 151093
      30 ..... ArgPr 31
151092 CTTCCTGTGCTATGAATCTGAGATTCCTGCTATGCTGTTCACAGGCC 151043
      31 oArGThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 47
      |||||
151042 CAGAGCCATGAGCTATCTCTTCCTTCCTGCAACGTCCCTGGTGGCTG 150993
      47 aLcysGlnValThrSerThrProAsnLysGlnLusnPhelYsThrCys 63
```

```
|||||
150992 TGTGCCAGGTAACTGCAGCCACAGCAACAGAACTTAAAAACATGT 150943
      64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
      |||||
150942 GCTGAGCTGGTTGAGAGGCTGCAGACTGGGTCTGCTCCGTTCCT 150893
      80 uProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 97
      |||||
150892 GCTGAGGCAATTGACTTCATTGCAGCGGACCTGCAGAGAGCTACACC 150843
      97 euSerGluProLeuAsnGlyAspLeuGlnGlyGlnTrpSerGlnLeuAla 113
      |||||
150842 TGTGTGAACCACTGGTGGGAACTTTTGGAGATATACACCACCTTGCC 150793
      113 ..... 113
150792 AGGTATCAGGGAATATAGCAGAGAGGTAGAAATCTTGTGACAGTGT 150743
      113 ..... 113
150742 CCTGGGTTCCAGATATGAGGGTAGAGCCTTGAGAACTGATGAAGT 150693
      114 ..... ArgLcysGlyIleTrpLeuSerLeuGly 123
      |||||
150692 TGTCACTGCTCCCTTCCCCCAAGGAAATGTGACTCTGCTCTTGGGT 150643
      124 GlyPheHisGluArgGlyGlnAspTrpGluGlnAsnGlnLysIleTyrAs 140
      |||||
150642 GGTTCATGAGCGTGGCCAAAGACTGGGAGCAGACTGAGAAATCTACAA 150593
      140 nCysHisValLeuLeuAsnSerLysGly..... 149
      |||||
150592 TTGCACGTGCTGCTGAACAGCAAGGAGAGACTTATATACCTTTAGCC 150543
      149 ..... 149
150542 TGCCCTTCCATCTCTTCTACTAGATTCTCCAGAAATTTTCTCAAC 150493
      149 ..... 149
150492 TCTATTTCCTTGACCAAGGATTTAGGGGTGTCCTACTTCAGTTCTTA 150443
      149 ..... 149
      150 ..... Ser...ValValAlaSerT 155
150442 GCCTATTAACATATCTCTCTGGGAGAGTAGCAAGGCTTCTAGAACAA 150393
      150392 CCAAGACTGATATTCCTTCTTCTTACTGTAGGGGACAGTACTGGCCACTT 150343
      155 YrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 171
      |||||
150342 ACAGAGAGACACATCTGTGTACGTAGAGATTCAGAGGAGGCGCTATG 150293
      172 ArgGluSerAsnTrpThrLysProGlyGlyThrLeuGluProProVal 188
      |||||
150292 TGTGAAAGCAACTCTACATCCCTGGGCCCGAGTCTGTAGTCACTGTAC 150243
      188 sThrProAlaGlyLysValGlyLeu..... 196
      |||||
150242 CACACACAGCAGCAAGGTAGGATTTGAAAGATGAGGAGGAGGAAACAG 150193
      196 ..... 196
150192 GAATACTTTGAACAGTGCAGTAGAGATAGAAAGCCCTAGAGAGGGGTA 150143
      196 ..... 196
150142 ATGGAATATGACATGATGCTGTGACAAACAGACAGACAGACTACTAAG 150093
      197 ..... AlaIleCysTyrAspMetArg 203
      |||||
```

```

150092 TAGGCTGTTTTCATTCAGATTGGTCTAGCTGTCTATGATGATCGG 150043
204 PheProGluLeuSerLeuLysLeuAlaGlnAlaGluValLeuLeuTh 220
|||||
150042 TTCCCTGAACCTCTCTGTCATTTGGCTCAAGCTGGAGCAGAGATCTTAC 149993
220 rTyrProSerAlaPheGlySerValThrGlyProAlaHisTrpGluVal 236
|||||
149992 CTAATCCTTACGCTTTTGATTCATTTACAGCCGACGCCACTGGAGGTAA 149943
237 Leu..... 237
149942 GATGATGCTTTTAAACATAGGGCCTTTCTTAACCTCATCTTCC 149893
237 ..... 237
149892 CCTTGGCCCTACACAGTTAAATTCCTTCCCTTCCACTAATGGGAAAC 149843
237 ..... 237
149842 TCATTCCCGAGATATTCTCTCATGAAATAGTTAAATAGTAATCATTC 149793
237 ..... 237
149792 TAGGCAATTACCAAAATAGTCAAAATGGTATGTTGCTGTATATGCC 149743
237 ..... 237
149742 TCACCTGTACATTCCTCCACTATTGCTACATGTACTTAAGTACACACA 149693
238 ..... 238
149692 TCTCATGCCAGAGTGTCTGCTGGGCGCCGTGCTATGAAACCCAGTGCTA 149643
248 rValIleAlaAlaAlaGlnCysGlyArgHisHisGluThrAlaAlaSer 265
|||||
149642 TGTATGGGAGAGACACAGCTGTGAGCGCCACCATGAGAGAGCAAGTT 149593
265 YrGlyHisSerMetValAlaAspProTrpGlyThrValAlaAlaArgCys 281
|||||
149592 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGGCCCTG 149543
282 SerGlyGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeu 298
|||||
149542 TCTGAGGGGCGAGGCTGTGCTGGCCGCAATAGACCTCAACATATGTGG 149493
298 ngIlnMetArgGlnHisLeuProValPheGlnHisArgArgProAspLeu 315
|||||
149492 ACAGTGGCGCCGACACCTGCTGTTCAGCAGCCGACGAGCCGTGACCTCT 149443
315 YrGlySerLeuGlyHisProLeuSer 323
|||||
149442 ATGGCAATCTGGGTACCCACCTGTCT 149417
seq_name: gb-pr:AL591806
seq_documentation_block:
LOCUS AL591806 200822 bp DNA linear PRI 30-JAN-2002
DEFINITION Human DNA sequence from clone RP11-544M22 on chromosome 1, complete
sequence.
ACCESSION AL591806
VERSION AL591806.16 GI:18476709
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 200822)
AUTHORS Harrison, E.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

```

```

COMMENT
On Feb 1, 2002 this sequence version replaced gi:17902927.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP11-544M22 is from the library RPCT-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-544M22. The true
left end of clone RP11-137A12 is at 156538 in this sequence. The
true right end of clone RP11-381D2 is at 145015 in this sequence.
FEATURES
source
1..200822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-544M22"
/clone_lib="RPCT-11.2"
10395..10397
/misc-feature
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
44098..44107
/misc-feature
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
59585..59604
/misc-feature
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
90797..90883
/misc-feature
/note="Single clone region. Reads generated from a
transposon library derived from a single pUC clone.
Restriction digest data confirm the assembly."
90861..90883
/misc-feature
/note="Sequence from un1-directional dGTP big dye
terminator reads only."
BASE COUNT 55215 a 49052 c 46183 g 50372 t
ORIGIN
alignment_scores:
Quality: 1211.50 Length: 660
Ratio: 3.946 Gaps: 10
Percent Similarity: 46.515 Percent Identity: 41.667
alignment_block:
US-09-357-675c-22 x AL591806 ..
Align seg 1/1 to: AL591806 from: 1 to: 200822
2 LeuGlyPheIleThrArgProPheHisGln.....LeuLeuCysTrh 15
|||||
172789 CTGGGCTTCATCCACGAGCCTCTCTCACAGATTCCTCTCCCTTGTGTCTC 172838
15 rGlyTyrArgLeuLeuArgIleProValLeuCysTrhGlnPro..... 29
|||||

```



```
* 108600 111672: contig of 3073 bp in length
* 111773 111772: gap of unknown length
* 111773 114662: contig of 2890 bp in length
* 114663 114762: gap of unknown length
* 114763 116145: contig of 1383 bp in length
* 116146 116245: gap of unknown length
* 116246 118147: contig of 1902 bp in length
* 118148 118247: gap of unknown length
* 118248 120579: contig of 2332 bp in length
* 120580 120679: gap of unknown length
* 120680 122618: contig of 1939 bp in length
* 122619 122718: gap of unknown length
* 122719 124539: contig of 1821 bp in length
* 124540 124639: gap of unknown length
* 124640 126551: contig of 2012 bp in length
* 126552 126751: gap of unknown length
* 126752 128460: contig of 1709 bp in length
* 128461 128560: gap of unknown length
* 128561 130592: contig of 2032 bp in length
* 130593 130692: gap of unknown length
* 130693 132864: contig of 2172 bp in length
* 132865 132964: gap of unknown length
* 132965 134881: contig of 1917 bp in length
* 134882 136682: contig of 1701 bp in length
* 136683 136782: gap of unknown length
* 136783 139138: contig of 2356 bp in length
* 139139 139238: gap of unknown length
* 139239 140757: contig of 1519 bp in length
* 140758 140857: gap of unknown length
* 140858 142558: contig of 1701 bp in length
* 142559 142658: gap of unknown length
* 142659 144303: contig of 1645 bp in length
* 144304 144403: gap of unknown length
* 144404 145490: contig of 1087 bp in length
* 145491 145590: gap of unknown length
* 145591 147171: contig of 1581 bp in length
* 147172 147271: gap of unknown length
* 147272 148639: contig of 1368 bp in length
* 148640 148739: gap of unknown length
* 148740 150295: contig of 1556 bp in length
* 150296 150395: gap of unknown length
* 150396 151505: contig of 1110 bp in length
* 151506 151605: gap of unknown length
* 151606 152813: contig of 1208 bp in length
* 152814 152913: gap of unknown length
* 152914 155423: contig of 2510 bp in length
* 155424 155523: gap of unknown length
* 155524 156880: contig of 1357 bp in length
* 156881 156980: gap of unknown length
* 156981 158045: contig of 1065 bp in length
* 158046 158145: gap of unknown length
* 158146 159255: contig of 1110 bp in length
* 159256 159355: gap of unknown length
* 159356 160417: contig of 1062 bp in length
* 160418 160517: gap of unknown length
* 160518 161644: contig of 1127 bp in length
* 161645 161744: gap of unknown length
* 161745 163180: contig of 1436 bp in length
* 163181 163280: gap of unknown length
* 163281 164483: contig of 1203 bp in length
* 164484 164583: gap of unknown length
* 164584 166039: contig of 1456 bp in length
* 166040 166139: gap of unknown length
* 166140 167290: contig of 1151 bp in length
```

alignment_scores:

```
Quality: 1187.00      Length: 835
Ratio: 3.854          Gaps: 12
Percent Similarity: 36.886      Percent Identity: 35.569
```

alignment_block:

US-09-357-675C-22 x AC105589

```
Align seg 1/1 to: AC105589 from: 1 to: 178418
2 LeucylphelethernarigProProHisIcn.....LeuLeuCysTh 15
|||||
39095 CTGGGCTTCATCACCAGGCTTCCTCATCACCCTGCTGCTGTGAC 39144
15 rGLyTfArgrleuLeuArgIleProValLeuCysThrIcnPro..... 29
|||||
39145 CGGATCCGATTACCTCAATCTCAGTACTTGTACTCAGCCAGGTAAAC 39194
29 .....
39195 ATATTATTGTTCTTTGTCTAGCAGTAGTGTCTAATGCATGTTATTA 39244
29 .....
39245 CCCGAATAGTGAGCTGCATCCACCAACAAATTAAATCGTGTACTCCAG 39294
29 .....
39295 TAGCTTAGGGTTATTAAATGCTGTGGAAGAAATGAACAGTGTAGGG 39344
29 .....
39345 TTGGTGGACACACCTCTAATCCACAGCTGGGGAACAGGAGAGAT 39394
29 .....
39395 CAACTAGAGATTAAATGATCATCCCGCTACATAGAAAGTCTGCTGGGC 39444
29 .....
39445 TACAAAAGACCCGGCTCCAAAGGGAGGTAATCAGAAAAAAG 39494
29 .....
39495 TAAATTCCGGCTCCACCCCACTTGAGCTCCCAAGTTTCTCCGGGCA 39544
30 .....
39545 GCGGCTAGAAATCTGAGAAAACCTGCTGTGCATTCACAGGCCAGAGC 39594
33 rMet.....SerSerSerThrSerTrpIleuProleuValAlaValc 48
|||||
39595 CATGGCAGAGCTCATCAGCTCTCTGGAGCTGCCCTGGCTGCTGCT 39644
48 ySGInValThSerThrProAsnIySGInIyAsnPhelysThrCysAla 64
|||||
39645 GCCAGGTACATCAACACCAACCAAGCAGAGACTTTAAACGTCGTGCT 39694
65 GluIeuValIySGInIyAlaIyArgrleuIyAlaCysLeuAlaPheLeuPr 81
|||||
39695 GAGCGTTCACAGAGCTACACGAGCTAGGCGGTCTCCCTGGCTTCTGCC 39744
81 oGLuAlaPheAspPheIleAlaIyAsnProIaIyLthrIleuLeuLeuS 98
|||||
39745 TGAGCATTTTACCTTATTGACAGCAATCTGCCAGACATTAACCTGCT 39794
98 eRGLIProIeuAsnIyAspLeuIySGInIySerGIeuLeuAla... 113
|||||
39795 CCGAACCTGTGATGGGACCTTTTGGGCAATATACCAAGCTTCCAGG 39844
113 .....
39845 TACAAGGTAGGCTGGGGTGGGGGAGGGGTANATCCCTGTGGGCGAG 39894
113 .....
39895 TGTTCCTGATTCACCATAGAGGGGTAGAGTTGTCAGTGCTCCCTCCC 39944
114 ..Arg.GluCysGLyIleTrpIeuSerLeuIySGIyPheHisIyArgrI 129
|||||
```

```

39945 CTCAGGGAAATGTGAATCTGGCTGCTCTGGGTGGTTCACCAACAGCTGG 39994
129 yglNapTPRgluInsnaglLysIleTyraSncYHisValLeuLeuA 146
| ::::::::::::::::::::|
39995 CCCAGATATGGGAACAGATCCGAAAAATCTACATTTGTCTGGGCTTCTGA 40044
146 sn.SerLys..... 148
| ::::::::::::::::::::|
40045 ACAACCCAGGGTGAAGACATTCATCTTCACATTTGGCTTTCTTAAGGTC 40094
148 ..... 148
40095 ATCTAATCTCTCTCACTTAGAACCTTAACATTTTCTCACTGGCT 40144
148 ..... 148
40145 TTAGGGGTTTTTTTGTGTTTGTGAATGCTCTGTTCTAGCCTAAAA 40194
148 ..... 148
40195 TTACTCTTTGGAGAGAGCTAAGCTGACCTTCAGTCTGTACTGAGC 40244
149 ..... GlySerValValAlaSerTyraIglLysThiSL 160
40245 ATTTCTCTTACTTTTGGATCAGTACTGCGCAGTTACAGAGAACACATC 40294
160 euCysaSPyAlgluIleProglLysIngLyProMetaIrgIuSerAsnTy 176
| ::::::::::::::::::::|
40295 TATGTATGTAGAGATCCAGGTCAGGGGCTTATGAGAGAACACATCTAT 40344
177 ThrLysProglLysIleThrLeuIngProProValLysThiProAlaLy 193
| ::::::::::::::::::::|
40345 ACCATGCTGTGATAGCTCTTGAACACCTGTCAAGACACACAGAGCAAA 40394
193 sValGly..Leu..... 196
| ::::::::::::::::::::|
40395 GGTAGGAGTTTGAAGAAAGAAATAGATATGTTGAAAAAGTTGTAACATC 40444
196 ..... 196
40445 TCCCTTTGGTGTGATGATGAGTAGTATTCATGTTACGTTGACAG 40494
197 ..... AlaIle 198
40495 GACAGGAGATGATGAGTGGTGTCTCTCGTTTCAGGGTGGTCTAGCAATC 40544
199 CysTyraSPMeTaIrgPheProglIuLeuSerIleLysLeuAlaIle 215
| ::::::::::::::::::::|
40545 TGTATGACATCGGTTCCCTGAACCTTCTGTAAATGGGCTCAAGCTGG 40594
215 yAlaGlulIleuThrTyProSerAlaIlePheGlySerValThiGlyPro 232
| ::::::::::::::::::::|
40595 GCCAGAAATACCTTACTTATCTTCACCTTCACCTTTGATCTGTTACGGGCCAG 40644
232 IaHisTPrGlVal..... 236
| ::::::::::::::::::::|
40645 CCCACTGGAGGTAGAGAGAATGCCCTCTTCATACATATAGGCTCCTTAA 40694
236 ..... 236
40695 ATCTATATTTTAGTAGTGGGAGACAGATAGACAAACCTAAACTTAAGC 40744
236 ..... 236
40745 ACTTCAGGGGATAGCCACAGTTTCAAGCCAGCATAGTATGCTCGAGT 40794
236 ..... 236
40795 AAAACAGAAACAAATFAGAGCTGGCAGTGGGGGTGCACACCTTTAAT 40844
236 ..... 236
40845 CCCAGCCTCAGAGAGCAGAAAGCAAGTGTGATGTGATCAAGACAGCCT 40894

```

```

236 ..... 236
40895 GGTCTACAGAGTAGTTCCAGAGACGCCGGCTACACAGAGAAACGGTC 40944
236 ..... 236
40945 TCAAAAACAAACAAACAAACAAACAAAGTAACCCACAGACACCCAAA 40994
236 ..... 236
40995 ATAGTGTGATGATGGTGTGTGCTTTTATCTCCGACATGTAAGCAGA 41044
236 ..... 236
41045 GCGCCGACATCTCTGTGGGTCCAGGCCAACCAAGTTATACAAGAGAC 41094
236 ..... 236
41095 CTATATATAAAAAATACAAAAACAAACACACACCCCTCCCTCTACTTGA 41144
236 ..... 236
41145 AGATATTTCTTCCTCCGATCTGTGAGAGAAATTCGTTCTTAGATAG 41194
236 ..... 236
41195 TGTGAGTATGTTATAGTTAGTTACTTAATCTAGTTGATCATATAGTCAAT 41244
236 ..... 236
41245 GAGTAGATTGGCCACAGTGTCCGCTTCTGTCACTCGCCCTAATAATG 41294
237 ..... Leu..LeuArgAla 240
41295 CTGTATGTTCTTAGTAGACGACGCTTCATCTCAGGTGCTGTGGGGCC 41344
241 ArgAlaIleGluSerIleCysTyraValIleAlaAlaIleAlaIleCysGly 257
| ::::::::::::::::::::|
41345 CGCGCCATGTAGTCTGCTGCTGCTGATGATGACAGACACAGTGTGAGC 41394
257 gHisHisGluThiArgAlaSerTyrgLysIleSerMetValValAspProt 274
| ::::::::::::::::::::|
41395 CCACCATGAGACAAAGCAAGTTACGGCCACACATGTTGTTGACCCCT 41444
274 rPluIleThiValAlaIleArgCysSerIleGlyProglLysLeuCysLeuAla 290
| ::::::::::::::::::::|
41445 GGGGACAGTGTGGCCAGCTGTTCGAGGACACAGGCTTTGCTTGTCT 41494
291 ArgIleAspLeuHisPheLeuIngIleMetArgIleHisLeuProValPh 307
| ::::::::::::::::::::|
41495 CGAATGATGTTCCACTTTCGCAACAGATGCGCCAAACATCTGGCTGTGT 41544
307 eGlnHisArgArgProAspLeuTyrgLysSerLeuGlyHisProLeuSer 323
| ::::::::::::::::::::|
41545 CCAGCATTCGACAGACCTGACCTTATGACAGTCTC...CTGCCACTCTCC 41590
seq_name: gb_ov:AF284575
seq_documentation_block:
LOCUS AF284575 1214 bp mRNA linear VRT 22-JUL-2000
DEFINITION Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.
ACCESSION AF284575
VERSION AF284575.1 GI:9367117
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 1214)
REFERENCE
Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
Bieganski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.

```


TITLE Crystal structure of the worm Nitfhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers

JOURNAL CURT. BIOL. 10 (15), 907-917 (2000)

MEDLINE 20414396

REFERENCE 2 (bases 1 to 1214)

AUTHORS Pace, H.C., Hodawadekar, S.C., Draganescu, A., Huang, J., Blegowski, P., Pekarsky, Y., Croce, C.M. and Brenner, C.

TITLE Direct Submission

JOURNAL Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA

FEATURES

source Location/Qualifiers

1..1214

/organism="Xenopus laevis"

/db_xref="taxon:8355"

gene 1..1214

/gene="Nifl"

CDS 180..1046

/gene="Nifl"

/note="Nifl: similar to the Nit domains of Caenorhabditis elegans Nitfhit and Drosophila melanogaster Nitfhit"

/codon_start=1

/product="Nit protein 1"

/protein_id="AA87104.1"

/db_xref="GI:9367118"

/translation="MAGAHKPLIAVCOMSTSDKKNFATCSRLIREAGRRACWFLPEAFDYIGSIEERLSLAESLHPTIORYQLAREGIMLSLAGEHKGPMDDORTSNGHVVYDNGHLYSVYRKALHPDVLQNGYSLSESSITLPGALIRITTPAGTIGLGVCTDLKPEPSIALAQOQAEELTPPSAFITTLAHEVLLRLRALETQCYVAAAO TDRHNERSTSGAMVYDPWGLVIGQDEGTGYAEIDIPYMERVRDMVWRHRRRT DLYGKISFNKPD"

BASE COUNT 281 a 305 c 345 g 282 t 1 others

ORIGIN

ALIGNMENT SCORES:

Quality: 933.50 Length: 311

Ratio: 3.734 Gaps: 2

Percent Similarity: 80.386 Percent Identity: 55.305

Alignment block:

US-09-357-675c-22 x AF284575 ..

Align seg 1/1 to: AF284575 from: 1 to: 1214

10 HISGLINLEUCYSTHRYTYRARGLEU.....ARGILEPROVA 24

102 CACAGCGCCCTTGTC.....AGACTCTATAGACTGTGACTTTAAC 142

24 LLEUCYTHRGINPROARGPROARGTHMETSERSETHRSERTPG 41

143 CCTGTGTCGGATATRAGGGGTTAGAACCTGAGTGCATGCGTGTGCC 192

41 LULEUPROLEUALAVALAVALCYSGINVALTHSERTHPROASINLSGLN 57

193 ACAAGCCCTGATGTGCGGTGCGACATGACTCAACCTCTGATTAAGGAG 242

58 GLUASNPHELYSTHRCYSAAGLEUVALGNGINLAALAAARGLEUGL 74

243 AAGAAATTTCCACCGCTTCGCGGCTGATCCGAGGAGCTCGCGGCTGC 292

74 VALACSLLEUALAPHELEUPROGLUALAPHEASPPHELEALARGASNP 91

293 CCCTTGACATGGGTGTTCTGCCGGAACCTTTGACTATATCGGGGCAACA 342

91 ROALAGLUTHREULEUSERGLUPROLEUASNGLYASPLEULEUGLY 107

343 TTGAGAGAGACGCTGAGTCTGGCTGACTCTCATATGGGACACACATTAC 392

108 GLNTRYSERGLINLEUALAARGGLUCYSGLYLETTPLEUSERLEUGLY 124

393 CGTTACACCACTCGCACGAGGTGTGGCTGTGCTTCCCTCGGGGG 442

124 YPHENHISGLIARGGLYGLINASPTRPGLUINASNGLINLYSILETYRASC 141

443 ATTTTCATGAGAAAGAGACCAACTGGGACAGGACCAACGATTTCCATT 492

141 YSHISVALLEULEUASNSERLYSGLYSERVALVALASERTYRARGLYS 157

493 CTCACCTGGTGTGGACAAACACAGGCACATAGATGGTGTACCCGAG 542

158 THHISLEUCYSPVALGUILLEPROGLYGLNGLYPROMETARGUSE 174

543 GCTCACCTGTTTAGCTGACGACAGAAAGAGTGTCACTCAGAGAGAG 592

174 RASNTYRTHRYSPROGLYGLYTHREUGLUPROVALYTHSRPROA 191

593 CATTTCCACCTCCCGGACAGACGTTATTGCCCATCTACTCTCCAG 642

191 LAGLYSVALGLYLEUALALECYSTRYRASPHELAGPHEPROGLU 207

643 CAGAAAGATTGGCTGGGGCTGTGTACGACCTCCCTCCAGAAATTC 692

208 SERLEUYSLEUALAGINALAGLYLAGLULLEUHTHYRPROSERAL 224

693 TCCTTGGCTGTGCCCAACAGAGGAGAACTTCTCACTTACCTTGTGC 742

224 APHEGLYSERVALTHRYGLYPROALAHISTYRGLUVALLEULARGA 241

743 CTCACCCCTCAGTACTGTCTGCGACATTTGGAGGTGTCTGAGAGCC 792

241 RGLAILEGLUSERGLNCYSTYRVALLEALALAAGLINCYSGLYARG 257

793 GTCCCATAGAAACCCAGCTGTACGTAGTTCAGCGGCACAGACAGA 842

258 HISHSGLUTHRRARGALASERTYRGLYHISSEMETVALVALASPR 274

843 CACAAATGAGAGAGAGCTCATGTGACGCTATGTGTGGAGACCCGTG 892

274 PGLYTHVALVALAALARGCYSSERGLUGLYPROGLYLEUCYSLEUAL 291

893 GGGGCTGTGATTTGCCCATATCCACGAGAACAGACGAATATGTATG 942

291 RGLLEASPLEUHSAPHELEUGLINLEARGSLINHSLEUPROVALP 307

943 AGATTGACATTCCTCATATGAGAGGTGTGAGGCGGACATGCCGTGTG 992

308 GLNHISARGARPROASPLEUTYRGLYSERLEU 318

993 AGGCACCGCAGACTGATCTGTAGGAAATC 1025

seq_name: gb_in:AF069986

seq_documentation_block:

LOCUS AF069986 1385 bp mRNA linear INV 23-JUL-1998

DEFINITION Caenorhabditis elegans nitrlase and fragile histidine triad fusion protein Nitfhit (Niflhit) mRNA, complete cds.

ACCESSION AF069986

VERSION AF069986.1 GI:3228663

KEYWORDS

SOURCE

ORGANISM

Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE

1 (bases 1 to 1385)

Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tiliib, S., Draganescu, A., Wermuth, P., Rothman, D.H., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

Nitrlase and fhit homologs are encoded as fusion proteins in drosophila melanogaster and caenorhabditis elegans

Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

JOURNAL

MEDLINE

98337986

2 (bases 1 to 1385)

Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tiliib, S., Draganescu, A., Wermuth, P., Rothman, D., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

Direct Submission

JOURNAL Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
 Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

FEATURES
 Location/Qualifiers
 source 1..1385
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 gene 1..1385
 /gene="Nlfhlt"
 CDS 3..1325
 /gene="Nlfhlt"
 /codon_start=1
 /product="nitrilase and fragile histidine triad fusion
 protein Nlfhlt"
 /protein_id="AAC39136.1"
 /db_xref="GI:3228664"
 /translation="MLSTVPRRTMATGRHFVAVCOMTSNDLEKNPOAKNMTERRAGE
 KRCMEVLPCEPFEFLINKNEQIDLAMTDCDEMEKRELARKHNITLSGGIHNDRP
 SDAAHPWNTHLIDSDGVTRAEYTNKLHLEFLETPGKRLMESEFSKGTMTPEVDTP
 IGRGLSTICVDVPEPLSLMNRKGAQLSPSPALNGLMETLRLARALENDCY
 VVAAOTGAHNPKEKROSGHSMVDPNGAVVACSEKRVDMCFADILSYVDLEMOY
 FSHRSDLYTLHINEKSESGELKFAFENIPADHIFSTPSFVFLKRVLDGCHLV
 SPKRVYPRLLDLDADLEFIYAKKVOAMLEKHNHNTSTTICVOGSKDAGQVPHNH
 IHLPRRAGDFGNETYOKLASDKPERKPRNSNEDMAEAVVYRNLM"

BASE COUNT 399 a 314 c 326 g 346 t

ORIGIN

alignment_scores:
 Quality: 706.00 Length: 284
 Ratio: 3.284 Gaps: 2
 Percent Similarity: 75.704 Percent Identity: 48.592

alignment_block:
 US-09-357-675C-22 x AF069986 ..

Align seg 1/1 to: AF069986 from: 1 to: 1385

```

32 ArgThmEtsrSerSerTherTrrpLuleuProLeuAlaAlaValCy 48
|||||
24 CGAACAAATGGGAGCTGACGAGCATTT.....ATCGCAGTATG 61
|||||
46 SGLValTherThrProAsnLysGlnGluAsnPhelYsthrCysAlaG 65
|||||
62 TCAATATGCAAGCGACCAACGATCTGGAAGAAAGAACTTCAAGCCCGA 111
|||||
65 LuleuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhelPro 81
|||||
112 ATATGATCGAAGACCGGAGAAAGAAAGTGGAGATGTTTCTTACCA 161
|||||
82 GluAlaPhasPheIleAlaArgAsnProAlaGluThrLeuLeuLeuSe 98
|||||
162 GAATGCTTCATTCATGGGCTCAACAAATGAGCAATGATGTTGGC 211
|||||
98 rGluPProLeuAsnGlyAspLeuLeuGlyGlnTySerGlnLeuAlaArg 115
|||||
212 GATGGCTACGACGCTGTAATATATCGGAAATATCGGAGTTGGCGA 261
|||||
115 LucysGlyIleTrpLeuSerLeuGlyLysPheHisGluArgGlyLys 131
|||||
262 AACACAAATATTTGGCTTTCACCTAGAGGCTTCATCACACAG.....GAT 305
|||||
132 TrpGluGlnAsnGlnLysIleTyAsnCySHisValLeuLeuAsnSerLy 148
|||||
306 CCATCAGACGCTGCTCATCCATCGAACACCCATGATCATAGATTTCCGA 355
|||||
143 sGlySerValAlaAlaSerTyArgLysThrHisLeuCysAspValGlu 165
|||||
356 TGGAGTACATCGTGTGAGTACAAACACCTTCACCTTCATTTGGAGA 405
|||||
165 lerpGlyGlnGlyPrometArgLysSerAsnTyThrLysProGlyGly 181
|||||
406 TTCCTGGAAAAGTACGATTGATGGAGCGAGTTCAGCAAGCTGGCACT 455

```

```

182 ThrLeuGluProProValLysThrProAlaGlyLysValGlyLeuAla1 198
::: |||||
456 GAGATGATTCCTTCAGTTCAGACCCCTATCGGACGCTCGACATTCAT 505
|||||
198 eCysTyArgMetArgPheProGluLeuSerLeuLeuAlaGlnAlaG 215
|||||
506 TTGCTACGACGCTTCCTCCAGAGCTGCTTCGGAACCGTAACGGC 555
|||||
215 lValGluIleLeuThrTyProSerAlaPheGlySerValThrGlyPro 231
|||||
556 GCGCCCACTGCTTCATTTCCAGTGCATTCACACGTATACAGGTCTT 605
|||||
232 AlaHisTrpGluValLeuLeuArgAlaAlaGlnLeuSerGlnCysTy 248
|||||
606 GCTCACTGGGAACCTGCTGCTGCGCCGCGCAATCGAATACAGTCTA 655
|||||
248 rAlaIleAlaAlaGlnCysGlyArgHisHisGluThrArgAlaSer 265
|||||
656 TTGTGTGGCGCTGCTCAACCTGGGCGCCACAACTCGAAGACAGCT 705
|||||
265 yGlyHisSerMetValAlaAspProTrrpGlyThrValAlaArgCys 281
|||||
706 ATGGGCATTCGATGCTGTGATCCGTGGGCGCTGTTCGACACAGT 755
|||||
282 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeu 298
|||||
756 TCCGAGAGAGTTCACATGCTGTGCTGTAATCGACTGAGTACAGTGA 805
|||||
298 nGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAspLeu 315
::: |||||
806 CACTCTCTCGTAAGTCAACACAGTATTCGCCACAGACGCTCGATCT 855
|||||
315 yr 315
||
856 AC 857

```

seq_name: gb_in:AF069989

seq_documentation_block:

LOCUS AF069989 1521 bp mRNA linear INV 23-JUL-1998

DEFINITION Drosophila melanogaster nitrilase and fragile histidine triad

fusion protein Nlfhlt (Nlfhlt) mRNA, complete cds.

ACCESSION AF069989

VERSION AF069989.1 GI:3228669

KEYWORDS

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1521)

Pekarsky,Y., Campidello,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Werbuth,P., Rotman,J.H., Huebner,K.,

Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Nitrilase and fht homologs are encoded as fusion proteins in

Drosophila melanogaster and caenorhabditis elegans

Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

9837986

2 (bases 1 to 1521)

Pekarsky,Y., Campidello,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Werbuth,P., Rotman,J., Huebner,K.,

Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Direct Submission

Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson

Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

Location/Qualifiers

1..1521

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="3"

/map="61A"

1..1521

/gene="Nlfhlt"

Quality: 1482.00 Length: 326
Ratio: 4.827 Gaps: 2
Percent Similarity: 94.172 Percent Identity: 84.356

alignment_block:
US-09-357-675c-22 x AAZ46101 ..

Align seg 1/1 to: AAZ46101 from: 1 to: 1416

```
2 LeuGlyPheIleThrArgProProHisGln.....LeuLeucysth 15
|||||
111 CTGGGCTTCATCCAGCGGCTCTCTCAGATGCCCTGCTCTGTCCTG 160
15 rGlyTyrArgLeuLeuArgIleProValLeuCysThrGlnProArgPro 32
|||
161 TGGACATCCGATACCTCACTCACTCACTTGTGCTGCACGCCAGGCCA 210
32 rGlyMetSer...SerSerThrSerTrpGluLeuProLeuValAlaVal 47
|||
211 GAGCCATGGCTATCTCTCTCTCTCTGCGAAGTGCCTGCGTGGCTGTG 260
48 CysGlnValThrSerThrProAsnLysGlnLysAsnPheLysThrCysAl 64
|||
261 TGGCAGGTACATCGAGCGCCAGACAAAGAACACTTTAAACATGTGC 310
64 agLLeuValIleGlnIleAlaIleArgLeuGlyAlaCysLeuAlaPheLeu 81
|||||
311 TGGAGCTGTTGAGAGAGGCTGCCAGACTGGGCTGCTGCTTCTCTGC 360
81 roGluAlaPheAspPheIleAlaArgAsnProAlaIleThrLeuLeu 97
|||||
361 CTGAGCATTTGACTTCATTCAGCAGGAGCCCTGACAGAGCTAACACCTG 410
98 SerGluProLeuAsnGlyAspLeuGlyGlnTyrSerIleuAlaIle 114
|||||
411 TCTGAACCACTGGTGGGAAACTTTTGAAGAAATACACCCAGCTTGGCAG 460
114 gGluCysGlyIleTrpLeuSerLeuGlyGlyPheHisGlyArgGlyGln 131
|||||
461 GGAATGTGAGCTGTGCTGCTCTTGGTGGTTCATGAGCGTGGCCAG 510
511 ACTGGAGGAGCACTCAGAAAATCTACAAATGTGCAGCTGCTGAACAGC 560
131 sPTpGluGlnAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 147
|||||
511 ACTGGAGGAGCACTCAGAAAATCTACAAATGTGCAGCTGCTGAACAGC 560
148 LysGlySerIleValAlaSerTyrArgLysThrHisLeucysAspVal 164
|||||
561 AAAGGGGCACTAGTGGCCACTTACAGAAAGACACATCTGTGACGTAGA 610
164 vIleProGlyGlnGlyProMetArgLysSerAsnTyrThrLysProGly 181
|||||
611 GATTCAGAGGAGGAGGCTATGTGTGAAGCAACTCTACCATGCTGGCC 660
181 LysThrLeuGluProProValLysThrProAlaGlyLysValGlyLeuAla 197
|||
661 CCAGTCTTGTGAGTCACTGTGCAGACACACAGCAGGCAAGATTGTAGCT 710
198 IleCysTyrAspMetArgPheProGluLeuSerLeuLysLeuAlaGln 214
|||
711 GTCGTATACATGCGGTTCCCTGAACTCTCTGCAATGTGCTCAAGC 760
214 agLValIleGlnIleLeuThrTyrProSerAlaPheGlySerValThrGly 231
|||
761 TGGAGCAGAGATACCTTACCTTCTCAGCTTTGGATTCATTACAGGCC 810
231 roAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluSerGlnCys 247
|||||
811 CAGCCCACTGGAGGTGTGCTCGGGCCGCTGCTATCGAAACCCAGTGC 860
248 TyrValIleLeuAlaIleGlnCysGlyArgHisIleGluThrArgAla 264
|||
861 TATGTAGTGGCAGCAGCAGTGTGAGCGCCACCATGAGAAAGAGAGCAG 910
XX
```

```
264 rTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC 281
|||||
911 TATGCGCACAGCATGCTGTAGACCCCTGGGAGACAGTGTGGCCGCT 960
281 ySerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPheLeu 297
|||||
961 GCCTGAGGGGCGAGGCTCTGCTGCCCCGAAATGACCTCACTATCTG 1010
298 GlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAsp 314
|||
1011 CGACACTGTGCGCGACACCTGCTGTGTTCCAGACCGCAGGCTTACCT 1060
314 yTyrGlySerLeuGlyHisProLeuSer 323
|||||
1061 CTATGGCAATCTGCTGACCCACTGTCT 1088

seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAF16257
seq_documentation_block:
ID AAF16257 standard; cDNA; 1382 BP.
XX
AC AAF16257;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular;
vulnerable; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO20005174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587513/55.
XX
P-PSDB: AAB57054.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 1; Page 1124; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytoskeletal,
CC cardioactive, immunomodulatory, muscular, vulnerable, gastrolntestinal,
CC nephrotoxic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
```

50 Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

alignment_scores:

Quality: 1457.00 Length: 327
Ratio: 4.746 Gaps: 2
Percent Similarity: 93.884 Percent Identity: 84.098

alignment_block:

US-09-357-675c-22 x AAF16257 ..

Align seg 1/1 to: AAF16257 from: 1 to: 1382

```

1 MetLeuGlyPheIleThrArgProProHisGln.....LeuLeuGly 14
|||||
77 ATGCTGGGCTTCATCCACAGCCCTCCACAGATTCCTGTCCTTCGTG 126
14 sThrGlyTyrArgLeuLeuArgIleProValLeuGlyThrGlnProArg 31
|||||
127 TCCGAGACTCCGGATACCTCACTCACTGACTTGTGCTAGGCCAGGC 176
31 roArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 46
|||||
177 CCAAGAGCATGGCTATCTCTCTCTCTGCGAATGCCCTGGTGACT 226
47 ValCysGlnValThrSerThrProAsnLysGlnLysAsnPhelysThrCy 63
|||||
227 GTGGCCAGGTAACATCATGACGCCACAGCACAGACAGAACTTAAACATG 276
63 sAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 80
|||||
277 TGCAGACTGGTTGAGAGAGCTGCCACACTGGGCTGCTGCTGCTTCC 326
80 euProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 96
|||||
327 TGCCTGAGGATTTGACTTCATTCACAGCGACCTGCAGAGACGCTACAC 376
97 LeuSerLurProLeuAsnGlyAspLeuLeuGlyLntYrSerGlnLeuAl 113
|||||
377 CTGTCTGAACCATGCGGTGGGAAACTTTGGAAATACACCCAGCTTGC 426
113 aArgGluCysGlyIleThrLeuSerLeuGlyGlyPheHisGluArgLys 130
|||||
427 CAGGGAATGTGACTGTGGCTGCTGCTGCTGCTTCCATAGGCTGGCC 476
130 lAspTrpGluGlnAsnGlnLysIleTyrAsnCySHisValLeuLeuAsn 146
|||||
477 AAGACTGGAGACAGACTCAGAAATCTACATGTCTCAGCTGCTGTAAC 526
147 SerLysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspVa 163
|||||
527 AGCAAAAGGGGAGTAGTGGCCACTTACAGGAAGACATCTGTGTGAGCT 576
163 lGluIleProGlyLysGlnLysProMetArgGluSerAsnTyrThrLysPro 180
|||||
577 AGAATTCACAGGCGAGGCGCT.ATGTGTGAAGCAACCTTACCATGCTG 625
180 lLysGlyThrLeuGluProValLysThrProAlaGlyLysValGlyLeu 196
|||||
626 GGCCAGACTCTTGAAGTCACTGTACAGACACAGCAGGCAAGATTGCTCA 675
197 AlAlaIleGlyTyrAspMetArgPheProGluLeuSerLeuLysAlaGlu 213
|||||
676 GCTGTCTGTATGACATGCGGTTCCCTGAACCTCTCTGCGGATGGGCA 725
213 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerValThrG 230
|||||
726 AGCTGGAGCAGAGATCTTACTTACCTTCTTGTGCTTTGGATCCATTACAG 775
230 lProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluSerGln 246
|||||
776 GCCCAGGCCACTGGAGAGGTGTTGCTGCGGGCCGCTGATGAAACCCAG 825

```

```

247 CysTyrValIleAlaAlaGlnCysGlyTyrGlnHisGluThrArgAl 263
|||||
826 TGTATGTAGTGGCAGCAGCAGCAGTGTGGACGCCACCATGAGAGAGAGC 875
263 aSerTyrGlyHisSerMetValValaAspProTrpGlyThrValValAla 280
|||||
876 AAGTTATGGCCACAGCATGGTGGTGAACCCCTGGGGAACAGATGGTGCCC 925
280 rGlySerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuHisPhe 296
|||||
926 GCTGCTCTGAGGGGCGCAGGCTCTGCTGCTGCCCAATAGACCTCACTAT 975
297 LeuGlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAs 313
|||||
976 CTGGACAGAGTTGGCGCGACACACTGCTGTGTTCCAGCAGCAGAGCCCTGA 1025
313 pLeuTyrGlySerLeuGlnHisProLeuSer 323
|||||
1026 CCTCTATGGCAATCTGGGTGACCCACCTGTCT 1056

seq_name: /STDS1/gcgdata/hold-geneseg/genesegn-emb1/NA1999.DAT.AAX30398
seq_documentation_block:
ID AAX30398 standard; DNA; 1203 BP.
XX
AC AAX30398;
XX
DT 14-MAY-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS Homo sapiens.
XX
PN W0907891-A1.
XX
PD 18-FEB-1999.
XX
PE 04-AUG-1998; 98MO-US16235.
XX
PR 19-AUG-1997; 97US-0056732.
PR 05-AUG-1997; 97US-0054798.
PR 05-AUG-1997; 97US-0054803.
PR 05-AUG-1997; 97US-0054804.
PR 05-AUG-1997; 97US-0054806.
PR 05-AUG-1997; 97US-0054807.
PR 05-AUG-1997; 97US-0054808.
PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055386.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056365.
PR 19-AUG-1997; 97US-0056366.
PR 19-AUG-1997; 97US-0056570.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;

```

XX WPI: 1999-167452/14.
 DR P-PSDB: AAY10877.
 XX
 PT New isolated human genes encoding secreted polypeptides - useful for
 PS diagnosis and treatment of pathological diseases
 XX
 PS Claim 3: Page 265-266; 331pp; English.
 XX
 CC The specification describes secreted proteins and their corresponding
 CC polynucleotides which are useful for preventing, treating or ameliorating
 CC medical conditions, e.g. by protein or gene therapy. Pathological
 CC conditions can also be diagnosed by determining the amount of the
 CC secreted polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the products, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,
 CC prostate diseases, asthma, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
 CC thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 SQ Sequence 1203 BP: 292 A; 307 C; 325 G; 278 T; 1 other;

alignment_scores:
 Quality: 965.00 Length: 207
 Ratio: 4.874 Gaps: 0
 Percent Similarity: 95.652 Percent Identity: 87.440

alignment_block:

US-09-357-675c-22 x AAX30398 ..

Align seg 1/1 to: AAX30398 from: 1 to: 1203

117 GYLIETRPSEULENGLYGYPHEHISGLUARGLYGLNAPRRGI 133
 3 GGACTCTGGCTGCTTGGTGGTTCACAGCGTGCCAAAGACTGGA 52
 133 uGlnAsnGlnLysIleTyrAsnGlyHisValLeuLeuAsnSerIleGly 150
 53 GCAGACTCAGAAAATCTACAAATGTCACGTCTCTCTAAACAG.AAAGGG 101
 150 erValValAlaSerTyrArgLysThrHisLeuGlyAspValGluIlePro 166
 102 CAGTAGGGCCACTTACAGAGACACATCTGTGTGACGTAGAGATTCCA 151
 167 GLYGLIGLYPROMETARGLUSERASTYRTHLYSPRODLYGLYTHLE 183
 152 GGGCAGGGGCT.ATGTGTGAACCACTTACCACTGCTGGCCCAACTCT 200
 183 uGluProGluValLysThrProAlaGlyLysValGlyLeuAlaIleCysT 200
 201 TGAAGTACCTGTACGACACACAGCAAGATTGCTGCTGCTGCT 250
 200 yraspMetargpPheProGluLeuSerLeuAlaGlnAlaGlyAla 216
 251 ATGACATGGCGTCCCTGAACCTCTCTGTGCAATGGCTCAAGCTGGAGCA 300
 217 GlnIleLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaH 233
 301 GAGATACTTACCTATCCCTTACGCTTTGGATTCATTACAGCCAGCCCA 350
 233 sTRPGluValLeuLeuArgAlaArgAlaIleGluSerGlnCysTyrValI 250
 351 CTGGGAGGTTCCTGCGGGCCGCTGCTATCGAACCACCACTGCTAGTAG 400
 250 leAlaAlaAlaGlnCysGlyArgHisHisGluThrArgAlaSerTyrGly 266

401 TCGCAGACAGCAGTGTGACGCCACCATGAGAAAGACCACTTATGCG 450
 267 HisSerMetValValAspProTrrpGlyThrValValAlaArgCysSerG 283
 451 CACAGATGATGGTAGACCCCTGGGAAACAGTGTGCGCCGCTGCTGTGA 500
 283 uGlyProGlyLeuGlyLeuAlaArgIleAspLeuHisPheLeuGlnGlnm 300
 501 GGGGCCAGGCGCTGCTGCTCCCGAATAGACCTACACTTATCTCGACAG 550
 300 etArgGlnHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 316
 551 TCGCCGACACCTGCTGCTTCCACACCGCAGCGCTGCTATGCTGCG 600
 317 SerLeuGlyHisProLeuSer 323
 601 AATCTGGTCACTCCACTGCTT 621
 seq_name: /SIDSI/gogdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD25458

seq_documentation_block:

ID AAD25458 standard; CDNA: 1214 BP.

XX AAD25458;

DT 26-MAR-2002 (first entry)

XX Xenopus laevis Ntl1 cDNA.

XX Frog; Ntl1; cytosolic; neuroprotective; cellular pathway; therapy;
 KW apoptosis; proliferative disorder; degenerative disease; ss.

XX Xenopus laevis.

XX Key Location/Qualifiers

FT CDS 180..1046

FT /tag= a

XX /product= "Frog Ntl1 protein"

XX W0200187958-A2.

XX 22-NOV-2001.

XX 15-MAY-2001; 2001WO-US15664.

XX 16-MAY-2000; 2000US-204713P.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Croce C, Brenner C, Pekarski Y;

XX WPI: 2002-082984/11.

XX P-PSDB: AAE15792.

XX Isolated cDNA encoding human, mouse, frog and yeast Ntl2 proteins,
 PT useful to find molecules that mimic or antagonize Fhl1 interaction for
 PT the treatment of proliferative or degenerative diseases -

XX Claim 21: Page 60; 61pp; English.

XX The invention relates to isolated nucleic acids comprising a fully
 CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and
 CC mouse Ntl2 proteins. Ntl1 and Fhl1 proteins are encoded as fusion
 CC proteins in invertebrates and as separate polypeptides in vertebrates.
 CC Ntl1 and Fhl1 interact physically and functions in same cellular pathways.
 CC Molecules which bind Ntl2 and mimic or antagonise Fhl1 interaction are
 CC used to treat diseases in which activity of Ntl2 protein is altered in
 CC a mammal. Fhl1 mimics induce apoptosis and are particularly useful to
 CC treat proliferative disorders, whilst Fhl1 antagonists promote cell
 CC proliferation and are particularly useful to treat degenerative disease.
 CC The present sequence is frog Ntl1 cDNA.

50 Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

alignment_scores:

Quality: 933.50 Length: 311
Ratio: 3.734 Gaps: 2
Percent Similarity: 80.386 Percent Identity: 55.305

alignment_block:

US-09-357-675c-22 x AAD25458 ..

Align seg 1/1 to: AAD25458 from: 1 to: 1214

```

10 HisGlnleucysthrGlyTyrArgleu.....ArglleProva 24
11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 CACACGCCCTGTGTGC.....AGACGCTATAGACTGACTGACTTAAAC 142
24 lleucysThrGlnProArgProArgThrMetSerSerSerThrSertrpg 41
143 CCGTGTCCCGATATAGGGGTAGAACCTGATGCAATGGCTGTGCCCC 192
41 lueProleuValAlaValCysGlnValThrSerThrProAsnlysgln 57
193 ACAAGCCCCGATGCGCGTGCAGATGACTTCACCTCGATAGAGAG 242
58 GluAsnPhelyThrCysAlaGluLeuValGlnGluAlaAlaArgleuG1 74
243 AAGAAATTTCCACGCTGTCCGCGCTGATCCGGAGGCTCGGGCGCTCG 292
74 yAlaCysleuAlaPheleuProGluAlaPheAspPheIleAlaArgAsn 91
293 CGCTTGATGTGTCTTCCCGGAGGCTTTGACTATATGGGGGCGACA 342
91 roAlaGluThrleuLeuSerGluProleuAsnGlyAspleuGly 107
343 TTGAGAGACGCTGAGTGTGCTGACTCTACATGGGAGACACATTACAG 392
108 GlnTyrSerleuAlaArgGluCysGlyTyrleuSerleuGlyG1 124
393 CGTTACACCCCACTCGCAGAGTGGGCTGTGCTTCCCGGGGGG 442
124 yPheHisGluArgGlyGlnAspTrpGluGlnAsnGlyIleTyrAsn 141
443 ATTTGATGAGAAAGACCCCACTGGACAGGACCAAGCATTTCCAAATT 492
141 yshisValleuAsnSerlySeryValAlaAlaSerTyrArglys 157
493 CTCACGTTGTGGACACACAGGACACATGATGTGTGACGACAG 542
158 ThrHisleucysAspValGluIleProGlyGlnGlyProMetArgG1 174
543 GCTCACCTGTTGACGTAGACTTCAGAAATGAGTGCATCAGACAGAG 592
174 rAsnTyrThrlyProGlyThrleuGluProProVallyshThrPro 191
593 CAGTTCACCCCTCCCGGAGCAGAGCTTATTCGCCCATCACTTCCAG 642
191 laGlylyValGlyleuAlaIleCysTyrAspMetArgPheProGlu 207
643 CAGGAAGAATGGGCTGGGGGTGTGTACGACCTCCGCTTCCGAAATTC 692
208 SerleuLyseuAlaGlnAlaGlyAlaGluIleleuThrTyrProSer 224
693 TCCTTGCTGTGGCCCAAGAGAGAGAACTTCTCACTTACCTTGTGC 742
224 apheGlySerValThrGlyProAlaHisTrpGluValleuLeuArgla 241
743 CTTTACCCCTACTACTGTGTCTGGACATFTGGAGTGTGTGAGAGCC 792
241 rGaIaIleGluSerGlnCysTyrValIleAlaAlaIleCysGlyArg 257
793 GTGCGAATAGAAACCACTGCTAGCTAGTTGAGAGGACACAGACAGA 842

```

```

258 HisGlnThrArgAlaSerTyrGlyHisSerMetValAlaSprProtr 274
843 CACATGAGAGAGAGAGCTCTATGTCACGCTATGTGGGTAGACCCGTG 892
274 rGlyThrValAlaAlaArgCysSerGluGlyProGlyLeuCysleuAla 291
893 GGGGCTGTGATGTGGCCATGCCAGAGAGACAGAAATATGTATGCTG 942
291 rGlyleuAspleuHisPheleuGlnGlnMetArgGlnHisleuProval 307
943 AGATTGACATTCCTACATGAGAGCTGTGAGCGGACATGCGGTGTGC 992
308 GlnHisArgArgProAspleuTyrGlySerleu 318
993 AGGACCGCAGAGACTGATCTGTATGGAAATC 1025

```

seq_name: /SIDSI/gcdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABLI2225

seq_documentation_block:

ID ABLI2225 standard; cDNA; 1495 BP.

XX ABLI2225;

AC ABLI2225;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO2001/1042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

XX P-PSDB; ABB68122.

DR New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 31157; 21np + Sequence Listing; English.

PS The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA

CC sequences (ABLI01840-ABLI6175) and the encoded proteins

CC (ABBS737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;

alignment_scores:

Quality: 699.50 Length: 279
Ratio: 3.331 Gaps: 3
Percent Similarity: 75.269 Percent Identity: 50.179

alignment_block:

US-09-357-675C-22 x ABL12225 ..

Align seg 1/1 to: ABL12225 from: 1 to: 1495

```

45 ValAlaValAlcysGlnValThrSerThrProAsnLysGlnGluAsnPhely 61
   ::::::::::::::::::::
146 ATGCGTGGGTGATGATGCGCTCTACACGACAGACGCGCGCTAATCTTAG 195
   ::::::::::::::::::::
61 sTrnCysAlaGluLeuValGlnGlnAlaAlaArgLeuGlyAlaCysLeuA 78
   ::::::::::::::::::::
196 CCAAGTATAGACTAGTGTGATAGGGCCAAGTCACAGAACGCTGATGTC 245
   ::::::::::::::::::::
78 lApheLeuProGlnAlaPheAspPheIleAlaArgAsnProAlaGluThr 94
   ::::::::::::::::::::
246 TCTTCTGCTGAGTGTGCTGACTTTGTGGCGAGAGCGGACCCCAACT 295
   ::::::::::::::::::::
95 LeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyIntYrSerG1 111
   ::::::::::::::::::::
296 ATTGACCTCTCCGAGGGCTTGGAGCGGAGTAAAGCGCGCATACCGGGA 345
   ::::::::::::::::::::
111 nLeuAlaArgGluCysGlyIleThrPheLeuSerLeuGlyGlyPheHisGlu 128
   ::::::::::::::::::::
346 ATTGCGGAGTGCACACAAATTTGGATTTCCTGGGTGGCTGCACAGC 395
   ::::::::::::::::::::
128 rGlyGlnAspTrpGlnAsnGlnLysIleTyrAsnCysHisValLeu 144
   ::::::::::::::::::::
396 GG.....AACGATCAAAATCTTCAACGCTCATGTTTG 430
   ::::::::::::::::::::
145 LeuAsnSerLysGlySerValValAlaSerTyrArgLysThrHisLeu 161
   ::::::::::::::::::::
431 CTCACGAGAAAGGAGACTAGACAGCATATACAGAAAGCTGCACATGTT 480
   ::::::::::::::::::::
161 sAspValGluIleProGlyGlnGlyProMetArgGluSerAsnTyrThrL 178
   ::::::::::::::::::::
481 TGAATGTACAGCTAA...GAGGTTCGCTACGCGAATCGATCAAGTAA 527
   ::::::::::::::::::::
178 ysProGlyGlyThrLeuGluProValLysThrProAlaGlyLysVal 194
   ::::::::::::::::::::
528 CGCGGGGATCTGCTTGGAGCGCCAGTGCAGTCTCAGTTGGCCAGATA 577
   ::::::::::::::::::::
195 GlyLeuAlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysL 211
   ::::::::::::::::::::
578 GGGCTTCAGATTGCTACGACGCTGTTTGGCTGACCGCGGTGCTGCT 627
   ::::::::::::::::::::
211 uAlaGlnAlaGlyAlaGluIleLeuThrTrpProSerAlaPheGlySer 228
   ::::::::::::::::::::
628 CAGGAAGCTGGGTGCCAATTTGTTAACTATACCATCCGACATTCACAT 677
   ::::::::::::::::::::
228 alThrGlyProAlaHisTrpGlyValLeuLeuArgAlaArgAlaIleGlu 244
   ::::::::::::::::::::
678 CAACCGGTAAAGGCGCACTGGAAATCTCTCTGCGGCGCAGACGATAG 727
   ::::::::::::::::::::
245 SerGlnCysTyrValIleAlaAlaGlnCysGlyArgHisHisGluTh 261
   ::::::::::::::::::::
728 ACTCAATGCTTGTGGTGGCTGCGCTCAGATGAGTGGCAGCAACAGAA 777
   ::::::::::::::::::::
261 rArgAlaSerTyrGlnHisSerMetValValAspProTrpGlyThrVal 278
   ::::::::::::::::::::
778 GCGACAGAGTTGGGCGCAGCATGATGCTTACCCCTGGGGAACGTTAC 827
   ::::::::::::::::::::
278 AlaAlaArgCysSerGlnGlyPro...GlyLeuCysLeuAlaArgIleAsp 293
   ::::::::::::::::::::
828 TGGCTAGCTGACGACGAGACGAGACTGATATAGACAGCGCGAGGTGAG 877
   ::::::::::::::::::::
294 LeuHisPheLeuGlnImetaArgGlnHisLeuProValPheGlnHisAr 310
   ::::::::::::::::::::
878 CTTCGCTGCTCAATCTCTGTATCAGACCATGCGCTGCTGCAATATCG 927
   ::::::::::::::::::::
310 gArgProArgLeuTyrGlySerLeuGlnHisProLeu 322
   ::::::::::::::::::::
928 TCGAAAGACATCTACGCTTAAACGCTAACAACCTT 964

```

seq_name: /SIDSI/gcgsdata/hold-geneseq/gene-seq-emb1/NA2001B.DAT:ABL12224

seq_documentation_block:

ID ABL12224 standard; CDNA: 3548 BP.

XX ABL12224;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR P-P5DB; ABB68121.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Claim 1; SEQ ID NO 31154; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (ABB5737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WPI

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;

alignment_scores:

Quality: 699.50 Length: 279

Ratio: 3.331 Gaps: 3

Percent Similarity: 75.269 Percent Identity: 50.179

alignment_block:

US-09-357-675C-22 x ABL12224 ..

Align seg 1/1 to: ABL12224 from: 1 to: 3548

```

45 ValAlaValAlcysGlnValThrSerThrProAsnLysGlnGluAsnPhely 61
   ::::::::::::::::::::
1199 ATGCGTGGGTGATGATGCGCTCTACACGACAGACGCGCGCTAATCTTAG 1248
   ::::::::::::::::::::
61 sTrnCysAlaGluLeuValGlnGlnAlaAlaArgLeuGlyAlaCysLeuA 78
   ::::::::::::::::::::
1249 CCAAGTATAGACTAGTGTGATAGGGCCAAGTCACAGAACGCTGATGTC 1298
   ::::::::::::::::::::
78 lApheLeuProGlnAlaPheAspPheIleAlaArgAsnProAlaGluThr 94

```

```

|||||
1299 TCTTTGCGAGTGTGACTTTGGCGCAGAGCCGCAACCAACT 1348
95 LeuLeuSerGluProLeuAsnGlyAspLeuGlyIntyrSergI 111
::: |||||
1349 ATTGAGCTCTCGAGGCTTGACGGAGATTAAATGGCGCATCCGGA 1398
111 nLeuAlaArgLucysGlyIleTrpLeuSerLeuGlyLpheHisGlu 128
::: |||||
1399 ATTGGCGAGTGCACAAATTTGATTTCCCTGGGTGGCGTCGACGAGC 1448
128 rGlyGlnAspTrpGluGlnAsnGlyIleTyrAsnGlyHisValLeu 144
|||
1449 GG.....AACGATCAAAAATCTCAACGCTCACTTTTG 1463
145 LeuAsnSerLysGlySerValValAlaSerTyrArgLysThrHisLeu 161
|||||
1484 CTCACACGAAAGGGAACATAGCAGATACAGAAAGCTGCACATGTT 1533
161 SASpValGluIleProGlyGlnGlyPrometArgGluSerAspTyrTrl 178
|||||
1534 TGATGTTACGACTAA...GAGGTTCCGCTACGGCAATCAGATACAGTTA 1580
178 ySPrGlyGlyThrLeuGluProProValLysThrProAlaGlyVal 194
|||||
1581 CGCGGGATCTGCTTGAGCGCCAGTACACTCCAGTGGCGCAGATA 1630
195 GlyLeuAlaIleCysTyrAspMetArgPheProGluLeuSerLeuLys 211
|||||
1631 GGGCTCAGATTGCTAGCAGCTCGTTTCTGATCGCGCGGCTGCT 1680
211 uAlaGluAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySer 228
|
1681 CAGGAACTGGGTCCAAATTTGTTAACTACCATCCGATTCACATACG 1730
228 alThrGlyProAlaHisTrpGluValLeuArgAlaArgAlaIleGlu 244
::: |||||
1731 CAACGGTAAGCGGCGCTGAGGAAATCTCTCGGCGCCAGCCATAGAG 1780
245 SerGlnCysTyrValIleAlaAlaGlnCysGlyArgHisHisGluThr 261
|||||
1781 ACTCAATGCTTTGGTGGCTGCGCTCAGATAGTTGGCAGACACAGAA 1830
261 rArgAlaSerTyrGlyHisSerMetValAlaAspProTrpLysThrVal 278
|||
1831 GCGCAGAGTTGGGCGCAGACATATGTTAGCCCTGGGAAACCTAC 1880
278 alAlaArgCysSerGluGlyPro...GlyLeuCysLeuAlaArgLys 293
::: |||||
1881 TGCGTGAAGTGCAGCAGCAGAGCTGTATATAGCAGCGCCGAGGTGAC 1930
294 LeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPheGlnHisAr 310
|||
1931 CTTCCTGCTGCTCAATCTCTATATCAGACATGCCCTGCTTGAACATCG 1980
310 gArgProAspLeuTyrGlySerLeuGlnHisProLeu 322
|||
1981 TCGAAGCAGACTTCATGCTTAAACAGCTTACACCTT 2017
seq_name: /SDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL20222
seq_documentation_block:
ID ABL20222 standard; DNA; 5692 BP.
XX
AC ABL20222;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX

```

```

OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
23-MAR-2001; 2001WO-US09231.
XX
23-MAR-2000; 2000US-191637P.
XX
11-JUL-2000; 2000US-0614150.
XX
(PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
XX
DR
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12139; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB157737-AB17072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;

alignment_scores:
Quality: 699.50 Length: 279
Ratio: 3.331 Gaps: 3
Percent Similarity: 75.269 Percent Identity: 50.179

alignment_block:
US-09-357-675C-22 x ABL20222 ..
Align seg 1/1 to: ABL20222 from: 1 to: 5692

45 ValAlaValCysGlnValThrSerThrProAsnLysGlnGluAspPhe 61
::: |||||
2858 ATTGCTGGGTGATGATGCGCTTACAGCAGCAAGCGCGCTTAATCTTAG 2907
61 sThCysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeu 78
: |||||
2908 CCAAGTATATAGCTAGTATGATAGGCGCAAGTCAACAGAACGCTGCATGC 2957
78 lApeLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThr 94
|||||
2958 TCTTCTGCTGAGTGTGCTGCTTGTGGCGGAGCAGCGGCCCAACT 3007
95 LeuLeuSerGluProLeuAsnGlyAspLeuGlyIntyrSergI 111
::: |||||
3008 ATTGAGCTCTCGAGGCTTGACGGCAAGTTAAATGCGCATACCGGGA 3057
111 nLeuAlaArgLucysGlyIleTrpLeuSerLeuGlyLpheHisGlu 128
|||||
3058 ATTGGCGAAGTGCACAAATTTGATTTCCCTGGGTGGCGTCGACGAGC 3107
128 rGlyGlnAspTrpGluGlnAsnGlyIleTyrAsnGlyHisValLeu 144
|||
3108 GG.....AACGATCAAAAATCTCAACGCTCACTTTTG 3142
145 LeuAsnSerLysGlySerValValAlaSerTyrArgLysThrHisLeu 161

```


Ratio: 2.384 Gaps: 7
Percent Similarity: 68.858 Percent Identity: 38.062

Alignment_block:
US-09-357-675C-22 x AMD25457 ..

Align seg 1/1 to: AAD25457 from: 1 to: 1292

```

30 ArgProAlqYThMeSeSerSeThSerThriPruLeuProLeuVal 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 CGTGAAGAGCCATCTCT.....ACTYCCGCTGGCCCTCAATACA 74

46 aValaGlueValThrSerThProAsnLysGLInLaSPheLysThrc 63
   : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 GCTT...CAAGTTTCTTCATT.....AAATCAGATTAACCTTAACCGGG 115

63 ySalGLueValGLInGLuLaalAaLarGLueGLyAlaCysLeuAlaPhe 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 CTTGTAGCCCTAGTGGGGAGCAGCAAAAGCAAGCTGCCACATGTTCT 165

80 LeuProGLuAlaPheASPheLeIleLaarGLaSPheProLaGLuThrLeu 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166 CTGCTCGATGCTTCAAT.....TCTCATATGACACACACTTA 203

96 uLeu.....SerGLuProLeuAsnGLyASPheLeuGLyGLInTrs 110
   : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 CTTTCTGACTATGACAGAGAGATTCCCTGGAGACTCCACAAACAAAGCTT 253

110 eRGLLeuAlaArgGLuCysGLyLTrPLeuSerLeuGLyGLyPheHis 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 CTGAAGTGTCAAGAGAGACAGACATATATCTCATTTGAGAGCTCCATCC 303

127 GLuArGLyGLInaSPTrGLuGLaSPheInLysLTrPLeuAsnGLyHisVa 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 GAA.....GAGAGTCTGGGAAACTGATTAATACCTCTC 338

143 lLeuLeuAsnSerLysGLySerValAlaLalSerTyArgLysThrHis 160
   : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 TGTGTTGGGCTGTAGGAGATTACTGTGAACACAGAAAGATTCATC 388

160 eUCyASPAlaGLuLeuProGLyGLInGLyProMetLarGLuSerAsnTr 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 TGTGTGACATTAATGATCTCTGGAAATATTCCTTCAAGAAATCAACAA 438

177 ThrLysProGLyGLyThrLeuGLuProProValLysThrProLaGLy 193
   : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
439 TTGACCCCTGGTGAAGATTCTCC...ACATTGATAGCCCTTACGTCAA 485

193 sValGLyLeuAlaLTrECySTyrASPmetLarPheProGLuLeuSerLeu 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 AGTGGCCCTGGGCATCTGATGATATCCGCTTGGCGAGCTTGCACAAA 535

210 yLeuAlaGLaLalAaLysLLeuThrTyProSerAlaPheGLy 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
536 TCTATGCCAAAGAGGCTGCCACCTCTTGGTGATCTCGAGACTTCAAT 585

227 SerValThrGLyProAlaHisTrPLeuValLeuLeuAaLarGLaLal 243
   : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 CTGACCACAGAGCAAGCCCACTGGAGAGCTGTTACGAGCCGGGCTGT 635

243 eGLuSerGLInCysTyValLLeIleAlaLalAGInCysGLyArgHisHis 260
   : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 TGATATATCAGGTGATGTGGCTACAGGCTCTACGTCCGGGATGCACAA 685

260 luthArgAlaSerTyLysHisSerMetValAlaSPTrPLeuThr 276
   : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
686 CCTGTATGTGCTCGGGGACACACACACTGTGTGGATCTTGGGGGAG 735

277 ValValAlaArgCysSerGLyProGLyLeuCysLeuAlaArgLLeas 293

```

|||||
786 CCTGAACAGACTGGCTGAATTCGCGACAAATCCCATTTTAAACAGA 835
310 rgargProAspLeuTyr 315
::||| |||||
836 AACGACGACACCTCTAT 852

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA81479

seq_documentation_block:
ID AAA81479 standard: DNA; 69936 BP.

AC AAA81479;

DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

KM Meningococcus B; Menb; ds.

XX Neisseria meningitidis.

PN WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rapunoli R, Pizsa M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be
XX used in the diagnosis and treatment of N. meningitidis infection and
XX other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 547-567; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414

XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent

XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the

XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX AAA81453 represent Neisseria meningitidis MemB polynucleotide ORF

XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies

XX against them, can be used in the manufacture of a composition. The
XX composition can be used as a medicament (or in the manufacture of a

XX medicament) for treating, preventing or diagnosing infection due to
XX Neisserial bacteria. For example, some of the identified proteins could

XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neisseriae. Identification of sequences

XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious

XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX Multivalent vaccines have also been tried but none have successfully

XX overcome antigenic variability. The provision of further, complete
XX sequences may provide an opportunity to identify secreted or surface

XX exposed proteins that may be presumed targets for the immune system and
XX which are not antigenically variable or at least more conserved than

XX other more variable regions.

SO Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

alignment_scores:
Quality: 410.00 Length: 275
Ratio: 2.343 Gaps: 8
Percent Similarity: 63.636 Percent Identity: 37.455

alignment_block:
US-09-357-675C-22 x AAA81479/rev ..

Align seg 1/1 to reverse of: AAA81479 from: 1 to: 69936

45 ValAlaValGlySerGlnValThrSerThrProAsnGlnGluAsnPhely 61
|||||
27128 GTTCCGCGCTGCGAGATGTTGCGGCGTGTCCGCGGAAACCAACGTCCG 27079
61 sThrcysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeu 78
::: ||||| ::|||
27078 CGCATGAACGCGCTGTGCGACGGCGCGGACAGCGATTCGG 27029
78 lApheLeuProGluAlaPheAspPheIleAlaArgAsnProAlaGluThr 94
::: ||||| :: |||
27028 TGTGCTGCGCCGAAATATTGCGTGTGATGCGCGCAACGATACGCAAA 26979
95 LeuLeuSerGluProLeuAsnGlyAspLeuGlyGlnTyr.... 109
||| |||||
26978 CTGCGCGTGTGCGAGCCTTTGGCGCG.....GACCGTTTCACAC 26938
110SerGlnLeuAlaArgGlyCysGlyIleTyrLeuSerLeuGly 124
|||::: |||||
26937 GGCATTGACGCAACGCGGAAAGATGCGCGTGTGTG...TTGCGCG 26891
124 Ty.....PheHisGluArgGlyGlnAspTyrGluGlnAsn 135
|| |||
26890 GCACTGTGCGCGCTGCAAAAGCTGCGAGCGCGCT..... 26859
136 GlnIstIleTyrAsnCysHisValLeuLeuAsnSerIstGlySerValVa 152
::: |||
26858 ..AAAGTATGATATACCGCTGTGTGACGCGGCGGCAATTTTACGATG 26812
152 lAlaSerTyrArgIstThrHisLeuCysAspValGluIleProGlyGln 169
::: |||||
26811 GGGCGTGTACCAAAATGACACTCTCTGCGTTTCCGTTTGGCGAA. 26763
169 lYProMetArgIstSerAsnTyrThrIstProGlyGlyThrLeuGluPro 185
|||
26762 ..CGCTATGCGGAGCGCATACATCCGCGCGCGGCGGATGTG...CCG 26718
186 ProValIstThrProAlaGlyValGlyLeuAlaIleCysTyrAspMe 202
::: |||||
26717 CACTGTGCGGAGGAGCGGTGCGGCGGCGGCAATTTTACGATG 26668
202 lArgPheProGluLeuSerLeuValGlnAlaGluAlaGluIle 219
::|||
26667 CCGCTTCCCAATTTTCCGACCC.....CAGTGGCGTTTACGAT 26624
219 eUrThrProSerAlaPheGlySerValThrGlyProAlaHisTyrGlu 235
|||
26623 TGATGTGCGCGCGCGTGTACGCAACGCGGCGGCAAGCGCATTTGGAG 26574
236 ValLeuLeuArgAlaArgAlaIleGluSerGlnCysTyrValIleAla 252
::|||
26573 CTGCTGTGCGCGCGCGTGTGCGGAAACCAATTTTACGTTGCGCGC 26524
252 AlAGlnCysGlyArgHisIstThrArgAlaSerTyrIstHisSerM 269
||| |||||
26523 GGCACAGGCGCGTTTGCACGAAACGACGCGGCGCATTCGACACACA 26474
269 etValValAspProTyrPheValValAlaArgCysSerGluGlyPro 285
|||
26473 TGATGTGCGATCGTGGGCGACGTTGTGACGATTCGCCGAGGCGGAA 26424

[illegible]

```

50 ACCAATACTAGG... AACATCATGCGTTTGGCGCTGCTGCAGCTCAAGG 96
52 rthPrOaInLysGInGluAsnRheLysrThrCysAlaGluLeuValGInC 69
   :::: :::||| ::::
97 TTCAGAGACAAAGTGGCCAAATGTGTCCAAAGCGCGCTGCCAATAATAGG 146
69 LuAlaAlaLarg... LeuLylAlaCysLeuAlaRheLeuProGluAlaRhe 84
   |||: :::: ||| |||||: ::::
147 CGCGCGTCAAGAGCATTAACCCCGATTGTACTCTGCGCGAGTGGTTT 196
85 AspRhe..... lLeAlaArgAsnProAlaLutRheLeu 96
   ::: :::: |||: ::::
197 AATGCTGCATAGCGACCAACTACTTTCGGGAGTACGCGAGACCAATT.. 244
96 uLeuSerGluProLeuAsnGluAspRheLeuGInLutYrSerGluLeu 113
   :::: :::||| :::: ||| :::: |||
245 ..... CCGCATGGCTATTACCTCGCAACACCTCTCCAAATTGG 281
113 lArgGluCysGluLutRhrLeuSerLeuGInGluRheInGluArgGlu 129
   ||||: :::: :::: ||| :::: ||| |||
282 CAGAGACACACAGGTGTACATCTGGGCGGACAAATCTCTGAATGGGC 331
130 GlnAspTrpGluGlnAsnGlnLysLleTyrAsnCysHisLeuLeuAs 146
   :::|||: :::: |||||: :::: :::
332 ..... GAAACAGATGCATCTACACACCTCGACAGGTGGTGC 369
146 nSerLysGlySerValAlaLaserYrArgLysrThHisLeuCysAspY 163
   : :::: ||||: :::: ||||| :|||
370 ACCCACTGGTGCACCTTGGCGCAAGCATCCGCAATGCATCATCTTTGACA 419
163 aGluLutRProGluGlnLutProMetArgGluSerAsnYrThrLysPro 179
   :::: :||| :::: |||||: :|||
420 TGAATGTCAAGGTTGGCATTCGTTCAAGGATCGGAA... ACCGTGCC 466
180 GluYlThrLeuGluProProValLysrThrProAlaLysValGluLe 196
   :|||: :::: :::: ||||: ::::
467 GCAGGCATGATTCACCATCATCACTACAGTACGCGCAAGATGGCAT 516
196 uAlaLecYsTyrAspMetArgRheProGluLeuSerLeuLysLeuAlaG 213
   :::: |||||: :||| :|||: ::::
517 CGCGATTCGCTACGATTTTCGATTTCAGAGATGCGAGCGCTTATGCGA 566
213 lAlaGluAlaGluLutLeuThrYrProSerAlaRheGlySerValThr 229
   :|||: :|||: :|||: :|||: :|||: :|||
567 ACAGAGCGCTCGAGATGATCATCTACCGGTGCATTCAACATGACCACT 616
230 GluProAlaHisTrpGluValLeuLeuArgLalaArgLalaLecGluSerG 246
   ||||| :|||: :||| :|||: :||| :|||
617 GGTCCACTGCACCTGGGACCTATTGCGACGAGATCCGCTGCACATGACAA 666
246 nCysTyrValAlaAlaAlaAlaGlnCysGluYrArgHisHisGluThrArg 263
   : :|||: :|||: :||| :||| :||| :|||
667 ACTCTTTGGTGCACACATGCACCAACCGCGTGCACAAAGCGCGATATG 716
263 lAsrTrpGluHisSerMetValAlaAspTrpTrpGluThrValAlaAla 279
   :::: |||||: :|||: :|||: :|||: :|||
717 TAGCCATATGGCCATTCATGTGGTGGATTCATGGCGCAAGGTGCACAG 766
280 ArgCysSerGluGluProGluLysCysLeuAlaArgLutLeaRheHisRph 296
   :|||: :|||: :||| :|||: :|||
767 ACTGCCACTGAAGCGGAGAAATGTGTGGCCCATTAATTTTCTCCGA 816
296 eLeuGInGlnMetArgGlnHisLeuProValRheGlnHisArgArgPro 313
   :::: |||||: :|||: :|||: :|||
817 GGTGGAGGAGGTGGTGCAGAGATTCGCGTCTTGGCGCAAGAGCTGAG 866
313 sPLeuTyrGlySer 317
   |||||: :|||:
867 ATCTGTACGCCACC 880

```


OM of: US-09-357-675c-22 to: EST:* out_format : pfs
 Date: Apr 29, 2002 10:01 AM

About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=fastp.p2n.model -DEV=xlh
 -O=/cgn2.1/USPTO.spool/US09357675/rnatt_29042002_091756_3146/app.query.fasta.1.1208
 -DE=EST -OFMT=fastp -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELCP=6.000
 -DELXEXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdt
 -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=200000000 -USER=US09357675_EGNT_1.6497
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLIPXY -WAIT -THREADS=1

Search information block:

Query: US-09-357-675c-22
 Query Length: 323
 Database: EST:*
 Database sequences: 13736207
 Search time (sec): 2545.690000

score list:

Sequence	Strd	Ordt	ZScore	Escore	Len	Document
gb_hlc:AK004988	+	1726.00	3060.33	1.3e-161	2993	AK004988 Mus musculus adult ma
gb_hlc:AK002269	+	1705.00	3031.84	1.3e-159	1374	AK002269 Mus musculus adult ma
gb_est2:BI647584	+	1314.00	2331.77	6.5e-121	832	BI647584 603278164F1 NIH_MGC_1
gb_est2:BG636313	+	1302.00	2314.75	1.1e-119	869	BG636313 602828020F1 NCI_CGAP_K
gb_est1:AL5229152	+	1265.00	2248.76	5.2e-115	844	AL5229152 602985469F1 NCI_CGAP_1
gb_est2:BI332807	+	1260.50	2240.43	1.5e-115	863	BI332807 602985469F1 NCI_CGAP_1
gb_est2:BI32768	+	1224.50	2176.77	5.3e-112	801	BI327680 602978783F1 NCI_CGAP_1
gb_est2:AL520768	+	1221.00	2166.26	1.4e-111	830	AL520768 602105735F1 NCI_CGAP_1
gb_est2:BF780562	+	1169.00	2076.98	1.9e-106	821	BF780562 602105735F1 NCI_CGAP_K
gb_est2:BG636916	+	1151.00	2045.12	1.1e-104	793	BG636916 602488424F1 NIH_MGC_18
gb_est2:BG6358107	+	1140.00	2027.92	1.0e-103	638	BE855103 ux23906.y1 Soares thym
gb_est2:AL522844	+	1120.50	1988.60	1.6e-101	928	AL522844 602047063F1 NCI_CGAP_1
gb_est2:BF385518	+	1116.00	1982.44	3.6e-101	789	BF385518 602047063F1 NCI_CGAP_1
gb_est2:BF385518	+	1104.00	1959.47	6.8e-100	893	BF385518 DRABTC11 Rat DRG Libra
gb_est2:BF69604	+	1077.00	1911.70	5.6e-97	846	BF69604 603054985F1 NIH_MGC_12
gb_est1:AL520767	+	1075.50	1907.14	5.6e-97	992	AL520767 AL520767 LTI_NFL004.NE
gb_est2:BF780847	+	1061.50	1883.35	1.3e-95	966	BF780847 602105735F1 NCI_CGAP_K
gb_est2:BF37474	+	1060.00	1883.69	1.1e-95	686	BF37474 601780658F1 NCI_CGAP_1
gb_est2:BE287010	+	1057.00	1878.87	2.1e-95	654	BE287010 601097244F1 NCI_CGAP_1
gb_est2:BF780562	+	1048.50	1862.12	1.8e-94	744	BF780562 602105735F1 NCI_CGAP_K
gb_est1:AL522373	+	1047.50	1857.33	3.3e-94	960	AL522373 AL522373 LTI_NFL004.NE
gb_est1:BF322101	+	1034.00	1838.79	3.6e-93	593	BF322101 maa34802.y1 NCI_CGAP_1
gb_est1:AL559163	+	1031.00	1829.03	1.2e-92	861	AL559163 AL559163 LTI_NFL008.TC
gb_est2:BF6762506	+	1001.00	1776.65	1.0e-89	764	BF6762506 602733954F1 NIH_MGC_49
gb_est2:BF35664	+	1001.00	1774.21	1.4e-89	940	BF35664 601780828F1 NCI_CGAP_1
gb_est2:BF35664	+	999.00	1773.96	1.5e-89	708	BF35664 601780828F1 NCI_CGAP_1
gb_est2:BF711300	+	980.00	1766.25	1.4e-89	625	BF711300 1096112.x1 Human Insu
gb_est2:BF711300	+	980.00	1738.61	1.4e-87	790	BF703176 602686083F1 NIH_MGC_95
gb_est2:BF371411	+	954.00	1694.73	3.8e-85	726	BF371411 601223207F1 NCI_CGAP_1
gb_est2:BF371411	+	954.00	1692.99	4.7e-85	765	BF371500 601223207F1 NCI_CGAP_1
gb_est1:AL562604	+	945.00	1674.96	4.8e-84	853	AL562604 AL562604 LTI_NFL003.NE
gb_est2:BF157823	+	944.00	1671.95	7.0e-84	946	BF157823 603030404F1 NIH_MGC_11
gb_est2:BG94701	+	934.00	1659.18	3.6e-83	610	BG94701 602392071F1 NIH_MGC_94
gb_est2:BF626803	+	933.00	1657.94	4.2e-83	582	BF626803 un11f07.y2 Soares_mamm
gb_est1:AA086823	+	926.00	1646.19	1.9e-82	544	AA086823 m93967.y1 Soares
gb_est2:BF140308	+	918.00	1631.48	1.3e-81	561	BF140308 365533 MARC 2P1C Sus s
gb_est2:BF968273	+	917.50	1625.03	9.8e-81	900	BF968273 602266588F1 NIH_MGC_84
gb_est2:BF155483	+	911.50	1613.47	2.8e-81	813	BF155483 603027349F1 NIH_MGC_11
gb_est2:BF629497	+	902.00	1603.94	4.3e-80	509	BF629497 un38f09.y1 Soares_mamm
gb_est1:AL580840	-	901.00	1596.40	1.1e-79	830	AL580840 LTI_NFL008.TC

gb_est1:AL182190 + 900.00 1600.13 7.0e-80 519 i AL182190 uc64f01.r1 Soares_m
 gb_est1:AA209049 + 899.00 1597.91 9.3e-80 538 i AA209049 mw73h11.r1 Soares m
 gb_est1:AA396793 + 894.50 1589.89 2.6e-79 536 i AA396793 vb28h09.r1 Soares m
 gb_est2:BF752623 + 889.00 1576.14 1.5e-78 746 i BF752623 603028471F1 NIH_MGC
 gb_est2:BF306735 + 882.00 1566.88 5.0e-78 564 i BF306735 601104244F1 NCI_CGA

seq_name: gb_hlc:AK004988

seq documentation block:

LOCUS AK004988 2993 bp mRNA linear HTC 19-JAN-2002
 DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
 library, clone:J300013F05:nitrilase 1, full insert sequence.
 ACCESSION AK004988
 VERSION AK004988.1 GI:12836590
 KEYWORDS HTC: CAP trapper
 SOURCE Mus musculus (strain: C57BL/6J) adult male liver cDNA to mRNA,
 clone: J300013F05.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.

REFERENCE

1 (sites)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)

REFERENCE

2 (sites)
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE

3 (sites)
 Shibata, K., Itoh, M., Atzawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitsuwa, T., Tashiro, H., Itoh, M.,
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasliwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, Y., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichannel sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE

4 (sites)
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

REFERENCE

5 (bases 1 to 2993)
 Adachi, J., Atzawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arkawa, T., Baldarelli, R., Bono, H., Brownstein, M., Butt, C.,
 Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
 Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Horii, F.,
 Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S.,
 Kurihara, C., Matsuyama, T., Miyazaki, H., Nishi, K., Nomura, K.,
 Nunazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Schiml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
 Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, T.,
 Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
 Hayashizaki, Y.

TITLE

Direct Submission
 Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

LOCUS DEFINITION	1374 bp	mRNA	linear	HTC 19-JAN-2002
AK002269	Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610007B07.nltrilase 1, full insert sequence.			
ACCESSION	AK002269			
VERSION	AK002269.1	GI:12832130		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (strain:C57Bl/6J) adult male kidney cDNA to mRNA, clone:1b:RIKEN full-length enriched mouse cDNA library clone:0610007B07.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (sites)			
TITLE	Carninci, P. and Hayashizaki, Y.			
JOURNAL	High-efficiency full-length cDNA cloning			
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)			
PUBMED	99279253			
REFERENCE	10349636			
AUTHORS	2 (sites)			
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.			
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)			
PUBMED	20499374			
REFERENCE	11042159			
AUTHORS	3 (sites)			
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, D., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsui, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer			
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)			
MEDLINE	20530913			
PUBMED	11076861			
REFERENCE	4 (sites)			
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.			
TITLE	Functional annotation of a full-length mouse cDNA collection			
JOURNAL	Nature 409, 685-690 (2001)			
MEDLINE	5 (bases 1 to 1374)			
PUBMED	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Harada, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurikawa, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schiraldi, L., Shibata, K., Shibata, Y., Shihagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yatsushita, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.			
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to			

FEATURES
 source
 Location/Qualifiers
 1. 1374
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGD:MGI:1891931"
 /db_xref="taxon:10090"
 /clone="0610007B07"
 /sex="male"
 /tissue_type="kidney"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 1. 1374
 /gene="Nrl1"
 /note="data source:MGD, source key:MGI:1350916,
 evidence:ISS
 nrl1:lane 1"
 BASE COUNT 330 a 359 c 342 g 343 t
 ORIGIN
 Alignment_scores:
 Quality: 1705.00 Length: 323
 Ratio: 5.328 Gaps: 1
 Percent Similarity: 99.071 Percent Identity: 99.071
 alignment_block:
 US-09-357-675C-22 x AK002269 ..
 Align seg 1/1 to: AK002269 from: 1 to: 1374
 2 LeuglypheiIeThrArgProProHISglnLeuLeuCysThrGlyTyrar 18
 |||||
 119 CGGGGCTTATCATCAGGCGCTCCTCCACCAATCTGTGTACCGATACCG 168
 18 gtleuLeuATgIleProvalLeuCysThrgInProaGrProaGrThmets 35
 |||||
 169 ATTACTTCGAACCCACGTACTTGTGTACTGACGCCAGGCCAGAACCATGT 218
 35 erSerSerThrSerTPrpGluLeuProLeuValAlaValCysglnValrhr 51
 |||||
 219 CCTCATCAACTTCCTGGGAGCTGCCCTGGTGGCTGTGTGCCAGTAAACA 268
 52 SerThrProAsnLysGlnGlnAsnPhelysThrCysAlaGluLeuValGI 66
 269 TCACACACCAACACAGACAGAACTTTAAACAATGTGTGATGTGGTTGCA 318
 68 ngluaAlaIaArGLeuGlyAlaCysLeuAlaPheLeuProGluAlaPhe 85
 |||||
 319 AAGAGCTGCCAAGACTGGGTGCTTGCCTGGCCTTTCTTGCTGAGGCATTGG 368
 85 sPpHeIleAlaArGAsnProAlaGluThrLeuLeuLeuSerGluProLeu 101
 |||||
 369 ACTTTATTCACAGAAACCTGCCGAGACATTAATCTCTGTCCGAAACATG 418
 102 AsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAlaArgGluCys GlyI 118
 |||||
 419 AATGGGAGATCTTTGGGCCAATAATACCAAGCTTCCAGGAGGAATGTTGGA 468
 118 leTPrpLeuSerLeuGlyGlyPheHISGluArgGlyGlnAspTPrpGluGn 134
 |||||
 469 TCTGGCTGTCTTGGGCGGTTTCCACGAGCGTGGCCAAAGACTGGGAGCAG 518

135 AsnGlnIlyIleTyAsnCysHisValLeuLeuAnsSerLySgLySerVa 151
|||||
519 AATGAGAAATCTACAAATTTGTCATGCTCTTTGAACACAGGATCAGT 568
|||||
151 lValAlaSerTyArGlySthHisLeuCyAspValGluIleProGlyG 168
|||||
569 AGTGCCCGCTTACAGAGACACATCTGTGGATGTAGATCCAGATC 618
|||||
168 lnglyProMetArGluSerAnsTyThrLySProGlyGlyThrLeuGlu 184
|||||
619 AGGGGCCATGAGAGAAACAACTATACCAAGCCGAGGACCTTTGAG 668
|||||
185 ProProValLySthProAlaGlyLySValGlyLeuAlaIleCyStryAs 201
|||||
669 CCACCTGTCAAGACACCGCTGGCCAGGTGTGTACCAATCTGTATGCA 718
|||||
201 pMetArGheProGluLeuSerLeuLySLeuAlaGlnIleGlyAlaGluI 218
|||||
719 CATGGGTTCCCTGAACCTTCTTTGAAATTTGGCTCAAGCTGGGCAAAA 768
|||||
218 lLeuThrTyProSerAlaPheGlySerValThrGlyProAlaHisTrp 234
|||||
769 TACTTACTTATCTTACAGCTTTAGATCTGTACAGGTCCAGCCACTGG 818
|||||
235 GluValLeuLeuAlaArgAlaIleGluSerGlnCyStryValIleAl 251
|||||
819 GAGTGTCTGTGGGGCCCGCCCATTTGAATCTCAGTCTATGTAAATGAC 868
|||||
251 AlaAlaAlaGlnCySgLyArGHisHisGluThrArgAlaSerTyGlyHis 268
|||||
869 AGCAGCGGAGTGTGAGCCACCAATGAACAAAGACAAAGTTATGGCCATA 918
|||||
268 erMetValValAspProTrpGlyThrValValAlaArgCysSerGly 284
|||||
919 GCATGCTGTGACCCGCGGACACAGTGTGGCCCGCTGCTCGCAGGGA 968
|||||
285 ProGlyLeuCySLeuAlaArgIleAspLeuHisPheLeuGlnIleMetAr 301
|||||
969 CCAGGCCCTCTGCTGTGATGATGATCTCCACTTCTTACACAGATGCG 1018
|||||
301 gGlnHisLeuProValPheGlnHisArgProAspLeuTyGlySerL 318
|||||
1019 CCAACACCTGCTGTGTCCAGCACCGCAGACCTGATGCGCAGTGC 1068
|||||
318 euGlyHisProLeuSer 323
|||||
1069 TGGGTCACTCACTCTCT 1085
|||||
seq_name: gb_est2:BI647584
seq_documentation_block: 832 bp mRNA linear EST 12-SEP-2001
LOCUS BI647584 603278164F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5318631 5',
DEFINITION mRNA sequence.
ACCESSION BI647584
VERSION BI647584.1 GI:15561820
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 832)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov
Plate: L1807 row: 0 column: 16
High quality sequence stop: 810.
Location/Qualifiers
1. 832
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5318631"
/clone_1lb="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; site:1; NotI;
site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lohar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NCI_CGAP Library."
BASE COUNT 198 a 214 c 217 g 203 t
ORIGIN
alignment_scores:
Quality: 1314.00 Length: 263
Ratio: 5.073 Gaps: 1
Percent similarity: 98.479 Percent identity: 98.479
alignment_block:
US-09-357-675C-22 x BI647584 ..
Align seg 1/1 to: BI647584 from: 1 to: 832
1 MetLeuGlyPheIleThrArgProBromHisGlnLeuLeuCySthrGlyTy 17
|||||
48 ATGCGGGGCTTCATCACCAGGCGCTCCTCCACCAATCTGTGTACCGGATA 97
|||||
17 rArgLeuLeuArGlyIleProValLeuCySthrGlnProArGProArGThm 34
|||||
98 CGGATTAATTCGAATCCAGTACTTTGTACTCAGCCCAAGCCCAAGAACCA 147
|||||
34 eTserSerSerThrSerTrpGluLeuProLeuValAlaValCysGlnVal 50
|||||
148 TTCCTCATCAACTTCTGTGGAGCTGCCCTGTGTGCTGTGGCCAGTA 197
|||||
51 ThrSerThrProAsnLySgInGluAsnPhelySthrCysAlaGluLeuVa 67
|||||
198 ACATCAACACCAACACAGACAGAACTTTAAACATGTGTGAGTGGT 247
|||||
67 lGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaP 84
|||||
248 TCAGAGGCTGCCAGACTGGTGTCTGCTGCTGTGTGCTGAGGCAT 297
|||||
84 hAspPheIleAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100
|||||
298 TTGACTTTATTCACAGAAACCTGCCGAGACATTAATCTGTCGCCAACCA 347
|||||
101 LeuAsnGlyAspLeuLeuGlyGlnTySerGlnLeuAlaArgGluCysG 117
|||||
348 CTGAATGGGATCTTTGGCCCAATATACCAAGCTTGCAG. GAATGTGG 396
|||||
117 ylleTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluG 134
|||||
397 AATCTGGCTGTCTTGGCGGTTTCCACAGGCGTGGCCAAAGACTGGGAC 446
|||||
134 lAsnGlnIlyIleTyAsnCysHisValLeuLeuAnsSerLySgLySer 150
|||||
447 AATAATCAGAAATCTACAAATTTGTCATGCTCTTTGAACGCAAGGATCA 496
|||||
151 ValValAlaSerTyArGlySthHisLeuCyAspValGluIleProGly 167
|||||
497 GTAGTGGCCAGTTTACAGAGAAATACATCTGTGGATGTAGATCCACAG 546
|||||
167 yGlnGlyProMetArGluSerAnsTyThrLySProGlyGlyThrLeuG 184


```

|||||
547 TCAGGGGCCAATGAGAGAAAGCAACTATACCAAGCTGAGGCACCTCTTG 596
184 lnpProvallysthrProalaglylsValglyLeuAlaIleCysTyr 200
597 AGCCACTGTCAGACACCGGCTGGCAAGGTGGTCAGCAATCTGTAT 646
201 AspMetArpPheProgluLeuSerLeuLysLeuAlaGlnAlaGlyAlaG1 217
647 GACATCGGTCCTGACACTTCTTGAATTTGGCTACGCT.GGGGCAGA 695
217 uileLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaHisT 234
696 AATACTTACTTATCTTCACCTTGTGATCTGTACAGTCCAGCCCACT 745
234 rpluValleuLeu.ArgAlaArgAlaIleGluSerGlnCysTyrVal11 250
746 GGGAGGTGCTGCTGGCGGCCCGCCCATTTGAATCTCAGTCTATGTAAAT 795
250 eAlaAlaAlaGlnCysGlyArgHisGlnGluThrArg 262
796 AGCAGCAGCCGCACTGTGGACGCCACATTTGAACACAGG 832
seq_name: gb_est2:BG963013

```

seq_documentation_block:

```

LOCUS BG963013 869 bp mRNA linear EST 12-JUN-2001
DEFINITION 602828020F1 NCI_CGAP_CO24 Mus musculus cDNA clone IMAGE:4982813 5',

```

```

ACCESSION BG963013
VERSION BG963013.1 GI:14350650
KEYWORDS EST.
SOURCE house mouse.

```

ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 869)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LRAM10986 row: 9 column: 06
High quality sequence stop: 805.
Location/Qualifiers

```

FEATURES

source

```

1. 869
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4982813"
/clone_lib="NCI_CGAP_CO24"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 Kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 212 a 221 c 233 g 202 t 1 others
ORIGIN

```

alignment_scores:

```

Quality: 1302 00 Length: 276
Ratio: 4.932 Gaps: 2
Percent Similarity: 95.652 Percent Identity: 94.203

```

alignment_block:

```

US-09-357-675C-22 x BG963013
Align seq 1/1 to: BG963013 from: 1 to: 869
1 MetLeuGlyPheIleThrArgProProHisGlnLeuLeuCysThrGlyTyr 17
47 ATGCTGGGCTTCATCACCAGGCTCTCCACCACTCTGTGTACCGGATA 96
17 rArgLeuLeuArgIleProValLeuCysThrGlnProArgProArgTyrM 34
97 CCGATTACTTCTGAATCCAGTACT.TGTACTACCCCGCCAGGCCAGAAACA 145
34 etSerSerSerThrSerTrpLysLeuProLeuValAlaValCysGlnVal 50
146 TGTCTCATCACTTCTCTGGAGCTGCCCTGTGGCTGTGTGTCCAGSTA 195
51 ThrSerThrProAsnLysGlnGluAsnPhelSerThrCysAlaGluLeu 67
196 ACATCAACACCAACAAGCAAGAGAACTTTAAACATGTGCTGAGTTGCT 245
67 lGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaP 84
246 TCAAGAGGCTGCGACACTGGGTGCTTGGCTGCTTCTTGCCTAGAGCAT 295
84 heAspPheIleAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100
296 TTGACTTTATTGCAAGAAACCTTCCGAGACATTAATCTCTGTCCGAACA 345
346 CTGAATGGGAGATCTTNTGGCCCAATATATAGCCAGCTTGCACAGGAATG 395
101 LeuAsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAlaArgGluCysG1 117
117 yIleTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGlu 134
396 AATGTGGCTGTCTTGGCGGCTTTCACAGACGGTGGCCAAACATGGGAGC 445
134 lAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150
446 AGAATCAGAAATATCTACATTTGTATGTCTTTTGAACAGCAAGGATCA 495
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluIleProG1 167
496 GTAGTGGCCAGTTACAGAAACACACATCTGTGCATGTAGAGATCCACAG 545
167 yGlnGlyProMetArgGluSerAsnTyrThrLysProGlyGlyThrLeuG 184
546 TCAGGGGCGCAATGAGAAACCACTATACCAAGCTTGAGGCACTCTTGG 595
184 lnpProvallysthrProalaglylsValglyLeuAlaIleCysTyr 200
596 AGCCACTGTCAGACACCGGCTGGCAAGGTTGGTCTAGCAATCTGTAT 645
201 AspMetArpPheProgluLeuSerLeuLysLeuAlaGlnAlaGlyAlaG1 217
646 GACATCGGTCCTGACACTTCTTGAATTTGGCTACGCT.GGGGCAGA 695
217 uileLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaHisT 234
696 AATACTTACTTATCTTCACCTTGTGATCTGTACAGTCCAGCCCACT 745
234 rpluValleuLeu.ArgAlaArgAlaIleGluSerGlnCysTyrVal11 250
746 GGGAGGTGCTGCTGGCGGCCCGCCCATTTGAATCTCAGTCTATGTAAAT 795
250 eAlaAlaAlaGlnCysGlyArgHisGlnGluThrArgAlaSerTyrGln 267
796 AGCAGCAGCGAGTGTGGAGCCCA.CATGAACAAGACAGCAAGTATATGCC 844
267 is.SerMetValValAspProTyr 274
845 ATAGACATGTGTGTGAACCGTGG 868
seq_name: gb_est1:AL529152

```

```

seq_documentation_block:
LOCUS      AL529152                844 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSDD004YM06 5
ACCESSION  AL529152
VERSION    AL529152.1 GI:12792645
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 844)
AUTHORS   Li, W. B., Gruber, C., Jesse, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.
FEATURES
     source
         1..844
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CSDD004YM06"
            /clone_1bp="LTI_NFL001_NBC4"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain. Vector: pCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed
            by Life Technologies. Contact: Feng Liang Life
            Technologies, a division of Invitrogen 9800 Medical Center
            Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
            8371 Email : filiang@litech.com URL :
            http://fulllength.invitrogen.com"

BASE COUNT      182 a      238 c      235 g      189 t
ORIGIN
alignment_scores:
    Quality: 1265.00      Length: 263
    Ratio: 5.000          Gaps: 0
    Percent Similarity: 96.198      Percent Identity: 87.833

alignment_block:
US-09-357-675c-22 x AL529152 ..

Align seg 1/1 to: AL529152 from: 1 to: 844

```

```

61 LysThrCysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysIle 77
|||||
1 AAAAAATGCTGAGCTGCTGAGAGGCTGCACACTGGCTGCTGCTGCT 50
77 uAlaPheLeuProGluAlaPheAspPheIleAlaArgAsnProAlaGluT 94
|||||
51 GCGTTTCCTGCTGAGGCACTTTGACTTCTTGCACGGGACCCCTGCGAGA 100
94 hrLeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyGlnTyrSer 110
|||||
101 CGCTACACCTGCTGTAACCACTGGGTGGGAACTTTGGAGAAATACACC 150
111 GlnLeuAlaArgGluCysGlyIleTrrPleuSerLeuGlyGlyPheHisG 127
|||||
151 CAGCTTGGCCAGGGAATGTGACTGTGCTGCTGCTGCTGCTTCCATGA 200
127 uArgGlyGlnAspTrrPglGlnGlnAsnGlnTyrIleTyrAsnCysHisVal 144
|||||
201 GCGTGGCCAGAGCTGGAGCAGACTCGAATAATCTACAAATTGTCACGTGC 250

```

```

144 euLeuAsnSerIysGlySerValAlaIaSerTyrArgIysThrHisLeu 160
|||||
251 TCCTGAACGCAAGAGGGGAGTAGTGCCACTTACAGGAAGACACTTCG 300
161 CysAspValGluIleProGlyGlnGlyPrometArgGluSerAsnTyrTh 177
|||||
301 TGTGACGTAGACATTCACAGGCGAGGCTTATGTGGAAAGCAACTCTAC 350
177 rLysProGlyGlyThrLeuGluProValIysThrProAlaGlyLys 194
|||||
351 CATGCCCTGGCCGACGCTTGAATCACTCTGACACACACAGGAGGAGAGA 400
194 aIGlyLeuAlaIleCysTyrAspMetArgPheProGluLeuSerLeu 210
|||||
401 TTGGCTACTGCTGCTGATGACATCGGGTCCCTGAACTCTCTTGCA 450
211 LeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 227
|||||
451 TTGGCTCAAGCTGGAGCAGACAGATCTTACCTTCCACGCTTTGGATC 500
227 rValThrGlyProAlaHisTrrPgluValLeuLeuAlaArgAlaIleG 244
|||||
501 CATTCACAGGCCAGCCGACCTGAGAGGCTTGGCTGGCGCCGCTATCG 550
244 luserGlnCysTyrValIleAlaAlaAlaGlnCysGlyTrrHisHisG 260
|||||
551 AAACCCAGTCTATGTAGTGAGGACAGCAGACAGTGTGAGCCACCTYAG 600
261 ThrArgAlaSerTyrGlyHisSerMetValAlaAspProTrrPgluTh 277
|||||
601 AAGAGGCAAGTTATGCGCCACAGCATGTGTGAGACCCCTGGGAGACAT 650
277 lValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAsp 294
|||||
651 GGGGCGCCGCTGCTGAGAGGCGCAGGCTGCTGCTGCTGCTGCTGCTG 700
294 euHisPheLeuGlnGlnMetArgGlnHisLeuProAlaPheGlnHisArg 310
|||||
701 TCACACTATGTGGACAGTGGCGGACACCTGCTGCTGCTGCTGCTGCTG 750
311 ArgProAspLeuTyrGlySerLeuGlyHisProLeuSer 323
|||||
751 AGGCTGACCTCTATGCGCAATCTGGGTCAACCCACTGTCT 789

seq_name: gb_est2:B1332807

seq_documentation_block:
LOCUS      B1332807                863 bp      mRNA      linear      EST 30-JUL-2001
DEFINITION 602985469P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5138032 5',
ACCESSION  B1332807
VERSION    B1332807.1 GI:15017464
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 863)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: L14M1339 row: j column: 17
            High quality sequence start: 22
            High quality sequence stop: 825.

```


460 AAGACTGGAGCAGACTCAGAAAATCTACAATGTGTCGTCGTGAAC 509
147 SerLysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspVal 163
510 ACCAAAGGGGACGTAAGTGGCCACTTACAGAAAGACATCTGTGTGACGT 559
163 TgLuLeuProGlyGlnGlyProMetArgLysSerAsnTyrThrLysProG 180
560 AGAATTCACAGGACAGGGCCCTATGTGTGAACCACTACCATTCCTG 609
180 TgLyLeuThrLeuGlnProProValLysThrProAlaGlyLysValGlyLeu 196
610 GGCCAGCTCTTGTAGTACCTGTGCACACACAGCAGCAAGATGTGCTTA 659
197 AlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysLeuAla 213
660 GCGTGTGTATGACATGCGGTCCTGTAACCTCTCTGTCATGTGCTCA 709
213 naIaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerValThrG 230
710 AGCTGAGACAGAGATACCTTACCTATCTTCAGCTTTGGATCCATTACAG 759
230 TgProAlaHisTArgLysValLeuLeuArgAlaArgAlaIleGlySerGln 246
760 GCCACGCCACTGGAGGTGTGCTGCGGGCCGCTGCTATCGAACCCAG 809
247 CysTyrValIleAlaAlaAlaGlnCysGlyArgHisGluThrArgAl 263
810 TGTATGATGTGTGGACAGCAGCAGTGTGGACGCCACCATGAGAAAGAGC 859
263 aserTyrGlyHisSerMetValValAspPro 273
860 AAGTTATGGCCACAGCATGTGGKAGACCC 890

seq_name: gb_est2:BF780562

seq_documentation_block:

LOCUS BF780562 821 bp mRNA linear EST_12-JAN-2001
DEFINITION 602103103P1 NCI_CGAP_Kid14 Mus musculus cDNA IMAGE:4221289

5', mRNA sequence.

ACCESSION BF780562
VERSION BF780562.1 GI:12085595

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 821)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHAM9806 row: e column: 02

High quality sequence stop: 713.

Location/Qualifiers

FEATURES

Source

1. 821
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4221289"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life

BASE COUNT 196 a 213 c 213 g 199 t
ORIGIN

alignment_scores:

Quality: 1169.00 Length: 253
Ratio: 4.733 Gaps: 4
Percent Similarity: 97.628 Percent Identity: 96.838

alignment_block:

US-09-357-675c-22 x BF780562

Align seg 1/1 to: BF780562 from: 1 to: 821

1 MetLeuGlyPheIleThrArgProProHisGlnLeuLeuCysThrGlyTyr 17
12 ATGCTGGGCTTCATCACACAGCCCTCCTCACCAACTCCTGTGTACCGGATA 61
17 TArgLeuLeuArgIleProValLeuCysThrGlnProArgProArgThrM 34
62 CCGATTACTTGGAAATCCCACTACT .GTACTCAGCCCGAGGCCAGAACCA 110
34 etSerSerSerThrSerTyrPgluLeuProLeuValAlaValCysGlnVal 50
111 TGTCTCATCACTTCTCTGGAGCTGCCCCGTGGTGTGTGTGTGTGTGTGT 160
51 ThrSerThrProAsnLysGlnGlnAsnPheLysThrCysAlaGluLeuVa 67
161 ACATCAACACCAACAAGCAGAGAACTTTAAACATGTCTGAGTGTGT 210
67 TgInGluAlaAlaArgLeuGlnLysLeuAlaPheLeuProGluAlaP 84
211 TCAAGAGGCTGCCAGACTGGGTCTTGTCTGCGCTTCTCTGAGGAT 260
84 heAspPheIleAlaArgAsnProAlaGlnThrLeuLeuLeuSerGluPro 100
261 TTGACTTATTATTCACGAAACCTGCGAGACATTACTCTGTCCGAACCA 310
101 LeuAsnGlyAspLeuLeuGlnGlyTyrSerGlnLeuAlaArgLysGly 117
311 CTGAATGGGATCT .TTGGGCCAATATATAGCAGCTTCCAG .GAATGTGG 358
117 TyleTrrPleuSerLeuGlnGlyPheHisGluArgGlyGlnAspTrrP 134
359 AATCTGGCTCTCTTGGCGGTTTCCAGAGCGTGGCCAAAGACTGGGAGC 408
134 TnsngLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150
409 AGAATCAGAAATCTACAAATTCATGTGCTTTTGAACACAGACAGGATCA 458
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluIleProG 167
459 GTAGTGCCAGTATACAGAGACACATCTGTGCATGTAGAGATCCAGG 508
167 TgInGlyPheMetArgLysSerAsnTyrThrLysProGlyGlyThrLeu 184
509 TCAAGGGCCACTGAGAGAAAGCAATATACCAAGCCGGAGGCACTTGG 558
184 TrrProProValLysThrPro .AlaGlyLysValGlyLeuAlaIleCys .T 200
559 AGCCACCTGTCAAGACACCGCTGCTGGCAAGTTGGTTACCAATCTGTGG 608
200 TrrAspMetArgPheProGluLeuSerLeuLysLeuAlaGlnAlaGlyAla 216
609 ATGACAGCGGTCCTCCGAACTTCTTGAATGGGTCAACGCGGGGCA 658
217 Gln .IleLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaH 233
659 GAAATTAATCTTACTTATCTTACCGCTTGTGATCTGTATACAGGTCCACCC 708
233 TstrPgluValLeuLeuArgAlaArgAlaIleGluSer .GlnCysTyrVa 249

seq_name: gb_est2:BG436916

seq_documentation_block:

LOCUS BG436916 793 bp mRNA linear EST 14-MAR-2001

DEFINITION 602488424F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4620892 5', mRNA sequence.

ACCESSION BG436916

VERSION BG436916.1 GI:13343422

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 793)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>
Plate: LNCM1382 row: 0 column: 05
High quality sequence stop: 767.

FEATURES

Location/Qualifiers

source 1..793

organism="Homo sapiens"

db_xref="taxon:9606"

clone="IMAGE:4620892"

/clone_lib="NIH_MGC_18"

/tissue_type="large cell carcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pGB7; site_1: XhoI; site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 181 a 210 c 230 g 172 t

ORIGIN

alignment_scores:

Quality: 1151.00 Length: 267

Ratio: 4.622 Gaps: 5

Percent Similarity: 93.258 Percent Identity: 83.895

alignment_block:

US-09-357-675c-22 x BG436916 ..

Align seg 1/1 to: BG436916 from: 1 to: 793

42 LeuprPLeuValAlaValCysGlnValThrSerThrProAsnLysGlnI 58

|||||

3 CTrGCCCTGTGGCTGTGTGCCAGGTAACATCGACGCCAGCAAGCAACA 52

|||||

58 uasnPhelYstnrcysAlaGluLeuValGlnGlnuAlaAlaArgLeuGlyA 75

|||||

53 GAACCTTTAAACATGCTGACGTGCTGAGAGGCTGCCACAGCTGGG 102

|||||

75 lacysLeuAlaPheLeuProGlnuAlaPheAspPheIleAlaArgAsnPro 91

|||||

103 CCTGCTGGCTTTCTCCCTGAGGCAATTGACTTATTCACAGGACCT 152

|||||

92 AlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyL 108

|||||

153 GCAGAGACGCTACCTGTCTGACACCTGGGTGGGAAACTTTTGAGAA 202

|||||

108 nTrSerGlnLeuAlaArgGluCysGlyIleTrpLeuSerLeuGlyGlyP 125

|||||

203 ATACACCCAGCTTGTCCAGGAAATGTGACTGTGGCTGCTTGGGGGCT 252

|||||

125 hehISGLuArgGlyGlnAspTrpGluGlnAsnGlnLysIleTrpAsnCy 141

|||||

253 TCATGAGCGGTGGCCAGAGCTGGAGCACACACAGAAATTCACAAATTT 302

|||||

142 HisValLeuLeuAsnSerLysGlySerValAlaSerTrpArgLysTh 158

|||||

303 CAGGTGCTGTGACACGACAAAGGGGAGTAGTGCCACTTACAGAGAGAC 352

|||||

158 rHisLeuCyAspValGluIleProGlyGlnGlyPromeTarGlySera 175

|||||

353 ACATCTGTGTGACGTAGAGATTCCAGGGGAGGCGCTATGTGTGAAGCA 402

|||||

175 snTrThrLysProGlyGlyThrLeuGluProProValLysThrProAla 191

|||||

403 ACTTACACATGCTGGCCAGCTGTGACTGACCTGTACAGCACACAGCA 452

|||||

192 GlyLysValGlyLeuAlaIleCysTrpAspMetArgPheProGluLeu 208

|||||

453 GGCAGATTGGTCTGACTGTGCTATGACATGCGGTCCCTGAACTCTC 502

|||||

208 rLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTrpPro.Serala 224

|||||

503 TCTGGATTGGCTCAAGCTGAGAGACAGATTACTTACTTCTTACAGCT 552

|||||

225 PheGlySerValThrGlyProAlaHisTrpGluValLeuLeuArgAla A 241

|||||

553 TTTGGATTCATTACAGGCCACCCACCTGGAGGTTGCTCGGGCCCC 602

|||||

241 rAlaIleGluSerGlnCysTrpValIleAlaAlaAlaGlnGlyArg 257

|||||

603 GGTGCTATCCAAACCCAGTCTATGATGAGTGCGACACACAGTGTGACCC 652

|||||

258 HSHISGLuThrArgAlaSerTrpGlyHisSerMetValValAsp.ProT 274

|||||

653 CACCATGACAGAGAGCAAGTATATGCCACAGCATGTGTGTAGACCCCT 702

|||||

274 rPGLy.ThrValValAlaArgCysSer.GluGlyProGlyLeuGlySLeuA 290

|||||

703 GGGGAACAGAGTGTGGCCCGCTCTTGTAGGGGCCAGGCTGTGCTTGC 752

|||||

290 laArgIleAspLeuHisPheLeuGlnGlnMetArgGlnHis 303

|||||

753 CCCGANTAGACCTCAAGTATCTGCGACAGTGTGCGCGGAAC 793

seq_name: gb_est2:BE853807

seq_documentation_block:

LOCUS BE853807 638 bp mRNA linear EST 27-SEP-2000

DEFINITION ux23906.y1 Soares_thymus_2NBMT Mus musculus cDNA clone IMAGE:3472378 5' similar to TR:088526 O88526 NITRILASE HOMOLOG 1. [1] mRNA sequence.

ACCESSION BE853807

VERSION BE853807.1 GI:10312377

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 638)

AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE Tumor Gene Index

Ratio: 4.293 Gaps: 6
Percent similarity: 88.475 Percent identity: 78.305

alignment block:
US-09-357-675C-22 x B1822844 ..

Align seg 1/1 to: B1822844 from: 1 to: 928

```
1 MetLeuGlyPheIleThrArgProHisGln.....LeuLeuCy 14
|||||.....
39 ATGCTGGGCTTCATCACCAGGCTCTCTCCACATTCCTCTCCCTGTG 88
14 sThrGlyTyrArgLeuLeuArgIleProValLeuCyThrGlnProArg 31
1 |||.....
89 TCCTGGACTCCCGATACCTCAACTCTCACTTGTCTGCTCAGCCAGCC 138
31 rArGThrMetSer...SerSerThrSerThrGlnLeuProLeuValAla 46
|||||.....
139 CCAGAGCCATGGCTATCTCTCTCTCTCTCTGCGAAGTCCCTGGTGCT 188
47 ValCysGlnValThrSerThrProAsnGlnGlnAsnPheLeuThrCy 63
|||||.....
189 GGTGTCACAGGTAACTACAGCCAGCAGCAAGCAAGAACTTTAAACATG 238
63 sAlaGluLeuValGlnGluAlaIleArgLeuGlyAlaCysLeuAlaPhe 80
|||||.....
239 TCCTGAGGCTGTCAGAGGCTGCGAGACTGGGCTGCTCTGCTGCTTCC 288
80 euProGlnAlaPheAspPheIleAlaArgAsnProAlaGluThrLeuLeu 96
|||||.....
289 TCCTGAGGCTGTCAGGCTGTCAGGCTGTCAGGCTGTCAGGCTGTCAGC 338
97 LeuSerGluProLeuAsnGlnLeuGlyLeuGlyLeuGlyLeuGlyLeu 113
|||||.....
339 CTGTCTGAACCACTGGTGGGAACTTTTGGAAAGATACACCAGCTTGC 388
113 aArgGluCysGlyIleThrPheSerLeuGlyLeuPheHisGluArgGly 130
|||||.....
389 CAGGAAATGTCGACTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 438
130 InAspThrGlnInAsnGlnLeuIleTyrAsnCyHisIleValLeuLeu 146
|||||.....
439 AAGACTGGAGCAGACTCAGAAATCTACAATTTGACCTGCTGCTGCAAC 488
147 SerTyrGlySerValValAlaSerTyrArgLysThrHisLeuCysAsp 163
|||||.....
489 ACCAAGGGGCGAGTAGTGGCCACTTACAGAGACACATCTGTGACCT 538
163 IGIuIleProGlyGlnGlyProMetArgGluSerAsnTyrThrLysPro 180
|||||.....
539 AGAGATTCCAGGCAAG..GGGCTTATGTGTGAAGCAACTCTACCATGCC 587
180 LylGlyThrLeuGluProProValLysThrProAlaGlyLysValGlyLeu 196
||.....
588 GGCCAGCTTGTAGTCACCTGTCAGGCCACAGCAGGCAAGATGGTCTTA 637
197 AlaIleCysTyrAspMetArgPheProGluLeuSerLeuLysAlaIle 213
|||||.....
638 GCTGCTGTATGACATGGGCTTCCCTGAACCTCTCTGCGATGGCTCA 687
213 nAlaGlyAlaGluIleLeuThr..TyrProSerAlaPheGlySer...Val 228
|||||.....
688 AGCTGGAGCAGAGATACCTTACCTTCTTCACTTTTGGATCCCATATA 737
229 ThrGlyProAlaHisThrGluValLeuArgAlaArgAlaIleGlyLeu 245
|||||.....
738 CAGGCGCCAGCCCACTGGAGGTGTGGCGGCGCCGCTATGGAAC 787
245 rGlnCysTyrVal..IleAlaAlaIleGlnCysGly..ArgHisIleGly 261
|||||.....
788 CCAGTGCATATGTTAGTGGCCAGCAGCACTGTGGAAGCCACCATGAGAA 837
261 rArgAlaSerTyrGlnLysSerMetValValAspProThrGlyThrVal 278
```

```
|||||.....
838 GAGAGC..ACTTATGGGACACAGCATGTGCTAAACCTGGGGACAGGGG 886
278 aAlaArgCysSerGluGlyProGlyLeu 287
::.....
887 CCCGTCTCTGAGGGCAGGCTGCTG 915
```

seq_name: gb_est2:BF385518

seq_documentation_block:

LOCUS BF385518 789 bp mRNA linear EST 27-NOV-2000
DEFINITION 602047063F1 NCL_CGAP_L19 Mus musculus cDNA IMAGE:4196559 5',
mRNA sequence.

ACCESSION BF385518
VERSION BF385518.1 GI:11366823

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.

TITLE 1 (bases 1 to 789)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs@email.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LM9533 row: n column: 16

High quality sequence stop: 754.

Location/Qualifiers

1..789

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4196559"

/clone_id="NCL_CGAP_L19"

/lab_host="DH10B (TL phage-resistant)"

/note="Organ: Liver; Vector: pCMV-Sp6R6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP Library."

BASE COUNT 182 a 206 c 208 g 193 t

ORIGIN

alignment_scores:

Quality: 1116.00 Length: 239

Ratio: 4.831 Gaps: 4

Percent similarity: 96.653 Percent identity: 92.469

alignment_block:

US-09-357-675C-22 x BF385518 ..

Align seg 1/1 to: BF385518 from: 1 to: 789

```
1 MetLeuGlyPheIleThrArgProHisGlnLeuLeuCysThrGlyTyr 17
|||||.....
83 ATGCTGGGCTTCATCACCAGGCTCTCTCCACATTCCTCTCCCTGTG 132
17 rArgLeuLeuArgIleProValLeuCysThrGlnProArgProArgTyr 34
|||||.....
133 CCGATTACTTCGAATCCAGTACT..TGTACTCAGGCCAGGCCAGAACCA 181
34 eTserSerSerThrSerThrGlnLeuProLeuValAlaValCysGlnVal 50
|||||.....
182 TCCTCTCATCAACTTCTGAGAGCTGCCCTGCTGCTGTGTGCGAGGTA 231
```



```

51 ThrSerThrProAsnLysGlnGluAsnPhenylsThrcysAlaGluLeuVa
|||||
232 ACATCAACACCAACAAGCAAGAACTTTAAAAACATGCTGATGCT 281
67 1GlnGluAlaAlaArgLeuGlyAlaCysLeuAlaAlaPheLeuProGluAla 84
|||||
282 TCAGAAGAGCTGCAGACTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 331
84 heAspPheIleAlaArgAsnProAlaGluThrLeuLeuLeuSerGluPro 100
|||||
332 TTGCTTTAATGCGAAGAACCTGCGCAGACATTAATCTGCGCAACCA 381
101 LeuAsnGlyAspLeuLeuGlyGlnTyrSerGlnLeuAlaArgGluCysG 117
|||||
382 CTGAATGGGATCTTTGGGCGCAATATAGCCAGCTTCCAGGGAATGTGG 431
117 yIleTyrPheLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTyrP 134
|||||
432 AATCTGCTGCTCTTGGCGGTTTCCAGAGCGTGGCCAGACTGGGAGC 481
134 LnsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlySer 150
|||||
482 AGATCAGAAATCTACAAATGTCATGCTTTGAACACAGCAAGGATCA 531
151 ValValAlaSerTyrArgLysThrHisLeuCysAspValGluIleProG 167
|||||
532 GTAGTGGCCAGTTACAGAGAACACATCTGCGATGTAGAGATCCACG 581
167 yGlnGlyProMetArgLysLeuSerAsnTyrThrLysProGlyGlyThrLeu 184
|||||
582 TCAGGGGCCAATGAGAGAAAGCAACATACCAAGCCTGGAGGCACTCTTG 631
184 LurProValLysThrProAlaGlyLysValGlyLeuAlaIleCysTyr 200
|||||
632 AGCCACTGTCAGACACCGCTGGCAAGGTGGCTTACAGATCTGTAT 681
201 AspMetArgPheProGluLeuSerLeuLysLeuAlaGluAlaGlyLag 217
|||||
682 GACATGCCGCTTCCCTGACACTTCTTGAATGTCAGAGCTGGGG...C 728
217 LuIleLeuThrTyrProSerAlaPheGlySerValThrGlyProAlaH 233
:::|||||
729 AGATATATACTATCTTCACGCTTG...GATCTGTACATGACGCCA 775
233 sTrpGluValLeu 237
|||||
776 CTGAGAGTGTCTGTG 788

seq_name: gb_est2:BG667049

seq_documentation_block:
LOCUS BG667049 893 bp mRNA linear EST 30-APR-2001
DEFINITION DRABTCl1 Rat DRG Library Rattus norvegicus cDNA clone DRABTCl1 5',
mRNA sequence.
ACCESSION BG667049
VERSION BG667049.1 GI:13888971
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 893)
Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,
Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and
Zhang,X.
Distinct gene expression profiles of rat dorsal root ganglion
induced by peripheral nerve axotomy
Unpublished (2001)
JOURNAL
COMMENT
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China

```

```

Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@cn.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@cn.ac.cn) or Han Zeguang (hanzg@cn.ac.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source 1. 893
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRABTCl1"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
BASE COUNT 218 a 213 c 262 g 196 t 4 others
ORIGIN

alignment_scores:
Quality: 1104.00 Length: 272
Ratio: 4.658 Gaps: 4
Percent Similarity: 87.132 Percent Identity: 82.721

alignment_block:
US-09-357-675c-22 x BG667049 ..
Align seg 1/1 to: BG667049 from: 1 to: 893

60 PheLysThrCysAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCys 76
|||||
1 TTTAAACGTGTGCTGAGCTGTTCAAGAGGCTACCAAGACTGGGCGCTTG 50
76 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsnProAlaG 93
|||||
51 CTTGGCCTTTTGCTGAGGCACTTGTGATTTATGACAGAAATCTGCGG 100
93 LurThrLeuLeuLeuSerGluProLeuAsnGlyAspLeuLeuGlyGlnTyr 109
|||||
101 AGACATTACTCTCTCCGAAACCACTGATGGGAGCTTTTGGCAATAT 150
110 SerGlnLeuAlaArgGluCysGlyIleTyrPheLeuSerLeuGlyPheH 126
|||||
151 AGCCAGCTTGGCCAGCAATGTGGAATCTGCTGCTGGGTGCTTCCA 200
126 sGluArgGlyGlnAspTyrPruGluGlnAsnGlnLysIleTyrAsnCysHisV 143
|||||
201 CGAGCGTGGCCAGAACTGGAGAGACTCAGAAATATACAAATGTCTATG 250
143 AlLeuLeuAsnSerLysGlySerValValAlaSerTyrArgLysThrHis 159
|||||
251 TGCCTTCGAAACAGCAAGGATCAGTAGGCGCAGTTACAGAGACACAT 300
160 LeuCysAspValGluIleProGluGlnGlyPheMetArgLysLeuSerAsn 176
|||||
301 CTATGTGATGTAGAGATCCAGGTCAGGCGCTATGAGAGAAAGCAACTA 350
176 rThrLysProGlyGlyThrLeuLeuProValLysThrProAlaGlyL 193
|||||
351 TACCATGCTGTGATAGCGCTTTGAACACCTGTCAAGACACAGAGCA 400
193 yValGlyLeuAlaIleCysTyrAspMetArgPheProGluLeuSerLeu 209
|||||
401 AGTTGGTCTAGCAATCTGTATGACATGCGGCTTCCCTGGAACCTTCTCTG 450
210 LysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 226

```

```

|||||
451 AATATGGCTCAAGCTGGGAGAAATCTTACTTATCTTCAAGCTTTGG 500
226 yserValThrGlyProAlaHisTrpGluValLeuAlaArgAlaArgAla 243
501 ATCTGTACCGGCCAGCCCACTGGAGGTGCTGTGGGCGGCCGCCA 550
243 IeGluSerGlnCysTyrValIleAlaAlaGlnCysGlyArgHisHis 259
551 TTGAGTCTCAGTGTACGTGATGACGAGACAGTGTGGAGCCACCAT 600
260 GluThrArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 276
601 GAGACAGAGCAAGATTACGGGACACACATGTGGTTGA..CCCTGGG 649
276 rValValAlaArgCysSerGlnGlyProGlyLeu..... 287
650 ATGGGGGAGCACTGTGC...GAGGGACCAAGCTTGGCTGCTCAATGATCC 696
288 .....CysLeuAlaArgIleAspLeuHisPheLeu 297
697 ACTTCTGAAGAATGGCCGAATGGCGGTGATCGGACATA.....CTA 740
298 GlnGlnMetArgGlnHisLeuProValPheGlnHisArgArgProAsp 314
741 TGGAGTCGGCGCTTACTCTGGGTGGAGACACCGC.....CT 781
314 uTyrGlySerLeuGly 319
782 CTAGGCTCGGCTTTGGG 797

seq_name: gb_est2:B1769604
seq_documentation_block:
LOCUS B1769604 846 bp mRNA linear EST_25-SEP-2001
DEFINITION 60305498SF1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',
mRNA sequence.
ACCESSION B1769604
VERSION B1769604.1 GI:15761182
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 846)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNI)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNI at:
http://image.llnl.gov
Plate: L14M11512 row: j column: 22
High quality sequence stop: 838.
Location/Qualifiers
1..846
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5204469"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SportS;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb. Insert size
range 1-3 kb. Library is normalized and enriched for

```

```

BASE COUNT 179 a 240 c 226 g 201 t
ORIGIN
alignment_scores:
Quality: 1077.00 Length: 264
Ratio: 4.414 Gaps: 4
Percent Similarity: 92.424 Percent Identity: 82.576
alignment_block:
US-09-357-675c-22 x B1769604 ..
Align seg 1/1 to: B1769604 from: 1 to: 846
1 MetLeuGlyPheIleThrArgProPheHisGln.....LeuLeu 14
|||||
59 ATGCTGGGCTTCATACACAGGCTCCTCACAGATTCTCTGCTTCTGTG 108
14 sThrGlyTyrArgLeuLeuArgIleProValLeuCysThrGlnProArg 31
|||||
109 TCCTGGACTCCGGATACCTCAACTCTCACTTCTTGTCTCAGCCAGGC 158
31 rArgThrMetSer...SerSerThrSerTrpGluLeuProLeuValAla 46
|||||
159 CCAGAGCCATGGGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 208
47 ValCysGlnValThrSerThrProAsnGlyGlnGluAsnPheLeuThr 63
|||||
209 GTGTGCCAGTAACATCAGCAGCCAGCAAGCAACACACTTTAAACATG 258
63 sAlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
|||||
259 TGTGTAGCTGTTCGAGAGGCTGCCAGACTGGGTGCTCTGCTTTC 308
80 euProGluAlaPheAspPheIleAlaArgAsnProAlaGluThrLeu 96
|||||
309 TGCCTGAGCATTTGACTTCATTTGCACGGGACCTGCAGAGCGTACAC 358
97 LeuSerGluProLeuAsnGlyAspLeuGlyGlnTyrSerGlnLeuAla 113
|||||
359 CTGTCTGAACCACTGGGTGGAACTTTTGAGAAATACACCACTTTC 408
113 aArgGlyCysGlyIleTrpLeuSerLeuGlyGlyPheHisGluArgGly 130
|||||
409 CAGGCAATGTGACTGTGCTCTCTGCTGCTGCTTCCATGAGCGTGGC 458
130 IAspTrpGluGlnAsnGlnIleTyrAsnCysHisValLeuLeuAsn 146
|||||
459 AAGACTGGGAGCAGACTCAGAAATCTACAAATGTACAGTGCAGTGAAC 508
147 SerTyrGlySerValValAlaSerTyrArgGlyThrHisLeuCysAsp 163
|||||
509 ACCAAAGGGGAGTAGTGGCCACTTACAGGAAGACACATCTGTGACGT 558
163 IeGluLeuProGlyGlnGlyProMetArgGluSerAsnTyrThrLeuPro 180
|||||
559 AAGATTTCCAGGCGAGGGGCTATGTGTAAAGCAACTTACCATGCCG 608
180 IeGlyThrLeuGluProProValIleTyrThrProAlaGlyValGlyLeu 196
|||||
609 GGCCAGCTTGTGATCAGCTGTACAGACACACAGGCAAGATTTGGTCA 658
197 AlaIle.CysTyrAspMetArgPheProGluLeuSerLeuValAlaG 213
|||||
659 GCTGTCACTTATGACATCGGTTCCCTCAACCTCTCTGCGCATTTGGC 708
213 IAlaGlyValGluIleLeuThrTyrProSerAlaPheGlySerValThr 229
|||||
709 AAGCTGACACAGAGTACTTACTTACTTCAAGCTTTTGGATTCATTACA 758

```

full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

```
230 GlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluSerG1 246
||||| ||||||||||||||| |||||||
759 GCCCAGC.CACTGGAGGTGTCTGCGGC.CGTGCTATCGAACCAC 806
|||||
246 nCysTyrValIleAlaAla.AlaGlnCysGlyArgHis 258
|||||
807 GTGCTATGTAGTGGCAGCAGCACACTGTGGACGCCAC 844
|||||
```

THIS PAGE BLANK (USPTO)


```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JM109 PAD445
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-25
```

```

alignment_scores:
Quality: 210.00      Length: 338
Ratio: 1.304         Gaps: 14
Percent Similarity: 47.633 Percent Identity: 24.556
```

```

alignment_block:
US-09-357-675C-22 x US-08-294-871A-25 ..
```

```

Align seg 1/1 to: US-08-294-871A-25 from: 1 to: 1785
```

```

46 AlaValCysGlnValThr.....SerThrProAs 55
164 GCGAGTGTGTTTCACAAAGCTTTCCGCGCGCTGGCGCGGACATCACCTAG 213
55 nLysGlnGluAsnPhelyThr..... 62
214 GAGGAGCAGAGGTTTCATGACACGTCAGATGATCTTCAGTGGGACAAC 263
63 .....Cys 63
264 AAGTTCGATCGCGCGCGGAGACACGGAACAGTGTCTGCTGT 313
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
314 CTCGACATGCTGACGAAGCCGCGAGCGCGGCGGAATTTCATTGCTCT 363
80 uProGlu.....AlaPheAspPheLeAla 89
364 CCGGCAATCGCGCTTACGACCTTTCCCGCGCTGGCATTTTCACCGAGC 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
414 AGCGCGAGCTGATAGCTTCTATGAGACCGAA..ATGCGCGCGCGG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGlyCysGlyLeuTrpLeuSerLe 122
461 GTCCCTGCACCTTTGAGAAAGCCCGGAACTCGGATCGCTCAACT 510
122 u.....GlyGlyPheHisGluArgGlyGln 131
511 GCGCTACGCTGACTGCTGCTGGAAGCGCGCTCAAGCGCTCC..... 553
131 sPTPrGluGlnAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 147
554 .....TTCAACACGCTCATTTTGGTGGATAAG 580
148 LysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspValGln 164
```

```

581 TCAGCGAAGATCGTCGGCAAGTATCGTAAGATCCATTTC..... 619
164 uLeuProGlyGln.....GlyProMetArg.....G 173
620 ...CCGGCTCACAGAGATAGAGAGCCACGGCGCTTCACAGATCTTG 665
173 LuSerAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
666 AAAAGCGTTATTTCCAGCGCGGCGATCTCGCGCTTCATGACG 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPhePro 206
716 GACGCGCGGAAATGGGATGTTCACTGCAACAGATCGCGCTGGCTGA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyValGluLeuThr.....T 221
766 AGCCTGCGCGGCTGATGGGCTTCAGGGCGCGGAGATCATCTCGCGGCT 815
221 YrProSerAlaPheGlySerValThrGlyProAlaHisTrpGluValLeu 237
816 ACAACAGCGCGGACCAACATCCACCGTTCCCGACGACGACCGTAGC 865
238 .....LeuArgAlaArgAlaIleGluSerGlnC 247
866 TCCTTCGACCATCTCTATCGATGAGCGCGGCTTATCAAGACGGGCG 915
247 sTyrValIleAlaAlaGlnCysGlyArgHisHisGluThrArgAla 264
916 CTGGTCCGCGCGCGCGGCAAGTGGGCGATG...GAGGAAATCGATTC 962
264 eTyrGlyHisSerMetValAlaAspProTrpGlyThrValAlaArg 280
963 TCCTGCGGCATCTGATCATCTGCGCGCGGCGGGAATGCTGCTTC 1012
281 CysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeuHis 296
1013 ACTGACGCGCTGAAGACGAGGTATCACCGCGCGGCTGATCGATGCG 1062
296 eLeuGlnIleMetArgGlnHisLeuProValPhe...GlnHisArgArg 312
1063 CTGCGGGAACCTCGTGAACACATCTTCACTTCAAGCAGCATCGTCA 1112
312 roAspLeuTyrGly 316
1113 CCCAGCATATGCT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-63
seq_documentation_block:
Sequence 63, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanda, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294.871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIORITY APPLICATION DATA:
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JMI09 PAD469
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-63

```

```

alignment_scores:
    Quality: 210.00      Length: 343
    Ratio: 1.304        Gaps: 14
    Percent Similarity: 46.939    Percent Identity: 23.907

```

```

alignment_block:
US-09-357-675C-22 x US-08-294-871A-63

```

```

Align seg 1/1 to: US-08-294-871A-63 from: 1 to: 1785
46 AlaValGlySerGlnValThr.....SerThrProAs 55
164 GCGATGTGTTTCACACACGTTTCCCGGCCGCTGGCGGACATCACCAG 213
55 nlySGlnGluAsnPhelYThr..... 62
214 GAAGACACAGAGTTTCATGACACGTCAATGATGACTTGCAATGGGACAC 263
63 .....Cys 63
264 AAGCTCCGATCGCGCGCGGACACGCGAAGAGTGTGCTGCTGCTT 313
64 AlaGluLeuValGlnGluAlaIaArgLeuGlyValAcylsLeuAlaPheLe 80
314 CTCGACATGCTGACGAAAGCCGAGCGGCGCCGAAATTCATGTCCTT 363
80 uProGlu.....AlaPheAspHeIleAla 89
364 CCGCGAACTCGCGCTTACGACCTTCCCGCGCTGGCATTTACACGACG 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
414 AGCCGAGCTGATAGCTTATGAGACCGAA...ATGCCCGCGCGGTG 460
106 LeuGlyGlnThrSerGlnLeuAlaArgGlyGlySerLeuThrLeuSerLe 122
461 GTCCGTCACCTCTTGGAGAGGCGCGAAGCTCGGATCGGCTTCACACT 510
122 u.....GlyGlyPheHisGluArgGlyGln 131
511 GGGCTACGCTGACCTGCTGCGAAGGCGCGTGAAGCTGC..... 553
131 sPTPrGluGlnAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 147
554 .....TTCAACAGCTGCATTTGGTGGATAG 580
148 LysGlySerValAlaIaSerTyrArgLysThrIleuLysCysAspValAl 164
581 TCAGGCAAGATCGTGGCAAGTCAATGTAAGATTCATTGG..... 619
164 uIleProGlyGln.....GlyProMetArg.....G 173
620 ....CCGGGTACAAAGAGTACGAGCGCTACCGCGCTTCAGACATCTTG 665
173 IuSerAsnTyrThrLysPProGlyThrLeuGluProProValLysThr 189
666 AAAAGCGTATTTCAGCGCGGCGCATCTCGGCTCCGCTGATGACGCTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
716 GACCCCGGAAATGGGATGTCATCGCAACGATCGCCCGTGGCCGCA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrThrPro 223
766 AGCTGGCGGTGATGGCTCAGAGGCGCGAGATTCATGTC..... 808
223 eAlaIlePheGlySerValThrGlyProAlaHisTPrGluValLeu..... 237
809 .....GGCGGCTACACACGCGGACCCACATCCCATTTGTTCCCGAG 850
238 .....LeuArgAlaArgAl 242
851 CACGACACCTGACGCTTCACACATCTCCTATGACATGACGCGCGGCTC 900
242 AlieGluSerLincystyryValIleAlaAlaIaGlnCysLysArgHisH 259
901 TTATCAGAACGGGCGCTGTGTCGCGCGCGGAGGAGGAGGATG...G 947
259 IsGluThrArgAlaSerTyrGlyHisSerMetValAlaAspProTPrGly 275
948 AGGAGAACTGATGCTGCTGCGGCATCTTCATGCTGCGCCGACGACCGG 997

```

```

276  тгВаVаlVаlаdаgуssеrGl...glPrоbUglеuSуlеаlаr 291
      :::|||||      :::      |||
998  GAAATGTCGCTCTCACTAGACGCTGGAAAGACGAGGTGATCAGCGCGC 1047
      ::|||||      ::::::::::|||:::      |||
291  gLеaSRlеuNlSPhеlеuGlnGlnMеtаrGlnlSleuPrоVаlPhe. 307
      ::|||||      ::::::::::|||:::      |||
1048  CGTGAATCTCGATGCGTGGCGGGAATCTGCGAACATCTTCACTCA 1097
      ::|||||      ::::::::::|||:::      |||
308  .GlnlSаtаrGPrоASleuTyrlу 316
      |||||||      ::::::::::|||:::      |||
1098  AGCAGATCTCGACGCCAGCATATGT 1126

seq_name: /cgn2_6/plodata/2/lna/6A.COMB.seq:US-08-876-398A-25

seq_documentation_block:
  Sequence 25, Application US/08876398A
  Patent No. 6083752
  GENERAL INFORMATION:
    APPLICANT: IKENKA, Yasuhiro
    APPLICANT: NANBA, Hirokazu
    APPLICANT: TAKANO, Masayuki
    APPLICANT: YAJIMA, Kazuyoshi
    APPLICANT: YAMADA, Yukio
    APPLICANT: TAKAHASHI, Satomi
    TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
    NUMBER OF SEQUENCES: 70
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: FOLEY & LARDNER
      STREET: 3000 K Street, N.W.
      CITY: Washington
      STATE: D.C.
      COUNTRY: U.S.A
      ZIP: 20007-5109
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patent Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/876,398A
      FILING DATE: 16-JUN-1997
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/211,641
        FILING DATE: 11-APR-1994
        PRIOR APPLICATION DATA:
          APPLICATION NUMBER: WO PCT/JP93/01101
          FILING DATE: 05-AUG-1993
          PRIOR APPLICATION DATA:
            APPLICATION NUMBER: JP 340078/1992
            FILING DATE: 21-DEC-1992
            PRIOR APPLICATION DATA:
              APPLICATION NUMBER: JP 212692/1992
              FILING DATE: 10-AUG-1992
              ATTORNEY/AGENT INFORMATION:
                NAME: Wegner, Harold C.
                REGISTRATION NUMBER: 25,258
                TELECOMMUNICATION INFORMATION:
                  TELEPHONE: (202) 672-5300
                  TELEFAX: (202) 672-5399
                INFORMATION FOR SEQ ID NO: 25:
                  SEQUENCE CHARACTERISTICS:
                    LENGTH: 1785 base pairs
                    TYPE: nucleic acid
                    STRANDEDNESS: double
                    TOPOLOGY: linear
                  MOLECULE TYPE: DNA (genomic)
                  ORIGINAL SOURCE:
                    ORGANISM: Escherichia coli
                    STRAIN: JM109 pAD445

```

```

: FEATURE: CDS
: NAME/KEY: join(233...1141)
: LOCATION:
US-08-876-398A-25

alignment_scores:
    Quality: 210.00      Length: 338
    Ratio: 1.304        Gaps: 14
    Percent Similarity: 47.633      Percent Identity: 24.556

alignment_block:
US-09-357-675C-22 x US-08-876-398A-25 ..

Align seg 1/1 to: US-08-876-398A-25 from: 1 to: 1785

46 AlavalCysGlnValThr.....SetThrProAs 55
|||||
164 GCGATGTTTCACAAAGCTTTTCCCGCGCCGCTGGCCGGACATCACACTAG 213
|||||
55 nLysGlnGlnAsnPhelysthr..... 62
|||||
214 GAAGGACACAGAGGTTTCATGACACGTCAGATGATACTTCGAGTGGGACAA 263
63 .....Cys 63
264 AAGTCCGATGCGCGCGGAGACACCGAACAAGTCTGTCGTCCTT 313
64 AlagLLeuValGlnGluAlaAlaArgLeuGluValaCysLeuAlaPhe 80
|||||
314 CTGACACTGCTGACAAAGCCGCGAGCGGGCGCGGAATTTATTTGTCCTT 363
80 uProGlu.....AlaPheAspHeiLeaIa 89
|||||
364 CCCCAGACTGCGCGCTTACAGACTTCTTCCGCGCTGGCAATTCACGACG 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
|||||
414 AGCCCGAGCTCGAATAGCTTATGAGACCGAA...ATGCCCGCCCGGTG 460
106 LeuGlyGlnItyrSerGlnLeuAlaArgGluCysGlyIleThrPheSerLe 122
461 GTCCGTCACACTCTTGAGAAAGCCGCGGAACTCGGAGATCGGCTTCAGT 510
122 u.....GlyGlyPheHisGluArgGlyGln 131
511 GGGCTACGCTGAACCTCGTCGTCGAAGCGCGGTCAAGCTCC..... 553
131 spTrpGluGlnAsnGlnIlystIleTyraAsnCysHisValLeuLeuAsn 147
554 .....TTCAACAGCTCCATTTTGGTGGATAG 580
148 LysGlySerValValAlaSerTyrArgIlyThrHisLeuCysAspValG 164
581 TCAGGCAAGATCGCGCGCAAGTACGTAAAGATCCATTG..... 619
164 uIleProGlyGln.....GlyProMetArg.....G 173
620 ....CCGGGTACACAGAGTAGAGGCGCTACCGGCGCTTCCAGCATCTTG 665
173 LuSerAsnTyrThrLysProGlyGlyThrLeuGluProProValIlyThr 189
666 AAAAGCGTTTTCGAGCGCGGCGCATCTCGGCTGCCGATATGACGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPhePro 206
716 GAGCGCGCAAAATGGGAGATGTCATCTGCAACGATCCCGCTGGCGTGA 765
206 uLeuSerLeuLeuAlaGlnAlaGlyIleValGluIleuThr.....T 221
766 AGCGTCGGCGGTATGGGCGCTCAAGGCGCGGACAGTCACTTCGCGCGGT 815
221 ytrProSerAlaPheGlySerValThrGluProAlaHisTrpGluValLeu 237

```



```

11  ...  ...  ...  ...  ...
816 ACAACAGCGCGACCAATCCACCTTCCCGACGACGACCACTACG 865
238 .....LeuArgAlaArgAlaIleGluSerGlnC 247
866 TCCTTCACCATCTCTATGATGACGAGCGGGCTTATGACGAGCGGGC 915
247 sTyrValIleAlaIleAlaGlnCysGlyArgHisGluThrArgAla 264
916 CTGGTCGCGCGCGCGGCAAGTGGCGATG...GAGGAGAACGTCATGC 962
264 eTyrGlnHisSerMetValIleAspProTyrGlyThrValIleAlaArg 280
963 TGCTCGGCGCACTCTGATGATGCGCGCGGACCGGGAATCGTCTCTC 1012
281 CysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeuHis 296
1013 ACTACGACGCTGAGACGAGAGTATCACCAGCGCGCTCATCTCATCG 1062
296 eLeuGlnGlnMetArgGlnHisLeuProValPhe...GlnHisArgArg 312
1063 CTGCGCGGAACTGCGTGAACACATCTTCAACTTCAGACGATCGTCAGC 1112
312 roAspLeuTyrGly 316
1113 CCCAGCACTATGCT 1126

```

seq_name: /cgn2.6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-63

seq_documentation_block:

Sequence 63, Application US/08876398A

Patent No. 6083752

GENERAL INFORMATION:

APPLICANT: IKENAKA, Yasuhiro

APPLICANT: NAKABA, Hirokazu

APPLICANT: TAKANO, Masayuki

APPLICANT: YAJIMA, Kazuyoshi

APPLICANT: YAMADA, Yukio

APPLICANT: TAKAHASHI, Satomi

TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/876,398A

FILING DATE: 16-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/211,641

FILING DATE: 11-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/JP93/01101

FILING DATE: 05-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 340078/1992

FILING DATE: 21-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 212692/1992

FILING DATE: 10-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wegner, Harold C.

REGISTRATION NUMBER: 25,258

```

REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD469
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-63

```

alignment_scores:

Quality	Ratio	Length	Gaps
210.00	1.304	343	14

Percent Similarity: 46.939 Percent Identity: 23.907

alignment_block:

US-09-357-675C-22 x US-08-876-398A-63 ..

Align seg 1/1 to: US-08-876-398A-63 from: 1 to: 1785

```

46 AlAlaValCysGlnValThr.....SerThrProAs 55
|||||
164 GCGATGTTTTCACACGTTTCCCGCGCTGGCGGACATCACCAG 213
|||||
55 nLysGlnGluAsnPhelyThr..... 62
|||||
214 GAAGAGCAGAGGTTTCATGACACGTCAGATGATCTTCAGTGGAGAAC 263
63 .....Cys 63
264 AAGTCGATCGCGCGCGGAGACAGCGAAGATCGTCTGTT 313
64 AlaGluLeuValGlnGluAlaIleArgLeuGlyAlaCysLeuAlaPheLe 80
|||||
314 CTCGACATGTCGACGAACCGCGAGCGGCGGGAATTCATTGTTCTT 363
80 uProGlu.....AlaPheAspPheIleAla 89
|||||
364 CCCGAACTCGCGCTTACGACCTTCTCCCGCGCTGCGATTTCACCGACG 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
|||||
414 AGCGGACGCTGATAGTTTATGAGACGAA...ATCGCGCGCGGCTG 460
106 LeuGlnGlnTyrSerGlnLeuAlaArgGlyCysGlyIleThrPheSerLe 122
461 GTCCGTCACCTTTTGAAGAGCGCGGAACTCGGAGTGGCTTCAACT 510
122 u.....GlyGlyPheHisGluArgGlyAla 131
|||||
511 GGGCTAGCGTCACTCGTCGTCGAAGCGCGCTCAAGCGTGC..... 553
554 .....TTCAACACGTCATTTGGTGGATGAAG 580
148 LysGlySerValIleAlaIleAlaSerTyrArgGlyThrHisLeuCysAspValG 164
581 TCAGCAAGATGTCGCGCAAGTATCGTAAGATCATTTG..... 619
164 uLleProGln.....GlyProMetArg.....G 173
|||||
620 ....CCGGTCAACAAGAGTACGAGGCTTACCGCGCTTCACGATCTTG 665

```

```

173   lueisrntfrrtlhsprogilylthrleuglrprovalylshtr 189
      ||:::|||||:::|||||||:::|||||:::|||||:::
666   AAAAGCCTTTTTCGACCGGGCCATCTCGGCTTCCGGTATTAGACGTC 715
190   ProAlaGlyLysValGlyLeuAlaIleCysTyrtAspMetArgPheProG1 206
      ||:::|||||:::|||||:::|||||:::|||||:::
716   GAGCGCGCGAAATGAATGGATGTTCATCTGCACAGCATGCCCTGGCCCTGA 765
206   uLeuSerLeuLysLeuAlaGlnAlaGlyIleuIleLeuThrtyrProS 223
      |:::|||||:::|||||:::|||||:::|||||:::
766   AGCGTGCGGGGTGAATGGGCTCACAGGGCGCGCAATATCATCTGC..... 808
223   eAlaPheGlySerValThrGlyProAlaHisTrpGluValLeu..... 237
809   .....GGCGGCTACACACGCGGACCACCAATCCCATTTGTTCCCCAG 850
238   .....LeuArgAlaValAl 242
851   CAGGACCCTGAGACGTCCTTCACCATCTCCATTGATGCATGACGCGGGCTC 900
242   aliegluseGlncsytyrValIleAlaIleAlaGlnCysGlyArgHisH 259
      ::::|||||:::|||||:::|||||:::|||||:::
901   TTATCGAAGAGGGGCGCTGTGTCGCGGCGCGGGGCAAGGTGGGCATG...G 947
259   IsgluThrArgAlaSerTyrglyHisserMetValAlaAspProtyrply 275
      :::::|||||:::|||||:::|||||:::|||||:::
948   AGAGAAGACTGCATCTGCTCGGCCACATCTCTGCATCTGTCGGCCACGCGGG 997
276   ThrValAlaAlaArgCysSergIu...GlyProGlyLeuCysLeuAlaAr 291
      :::::|||||:::|||||:::|||||:::|||||:::
998   GAATGCTGCTCTCACTACATGACGCTGGAAGACAGAGTACACGCGCGGC 1047
291   glieAspRleuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe. 307
      :::::|||||:::|||||:::|||||:::|||||:::
1048   CGTCGATCTGTCGATCTGTCGCGGGGAATGCGTGAACACATCTTCAACTTCA 1097
308   ..GlnHisArgArgProAspLeuTyrgLy 316
      ||:::|||||:::|||||:::|||||:::|||||:::
1098   AGCAGCATGCTGCACGCCACACATCATGTGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-5
seq_documentation_block:
? Sequence 5, Application US/08294871A
? Patent No. 58245522
? GENERAL INFORMATION:
? APPLICANT: Ikenaka, Yasuhiko
? APPLICANT: Namba, Hirokazu
? APPLICANT: Takano, Masayuki
? APPLICANT: Yajima, Kazuyoshi
? APPLICANT: Yamada, Yukio
? APPLICANT: Takahashi, Satomi
? APPLICANT: Okubo, Kazuma
? APPLICANT: Yamada, Kazuhiko
? APPLICANT: Hiratashi, Yoshiko
? TITLE OF INVENTION: Immobilized Enzyme Preparation and
? TITLE OF INVENTION: Process for Producing D-a-Amino Acid
? NUMBER OF SEQUENCES: 70
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Foley & Lardner
? STREET: 3000 K Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.
? ZIP: 20007-5109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/294.871A
```

```

1  FILING DATE: 22-AUG-1994
2  CLASSIFICATION: 435
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER: US 07/971,758
5  FILING DATE: 12-APR-1993
6  PRIOR APPLICATION DATA:
7  APPLICATION NUMBER: US 07/917,111
8  FILING DATE: 07-AUG-1992
9  PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 08/211,641
11 FILING DATE: 11-APR-1994
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: JP 400848/1990
14 FILING DATE: 07-DEC-1990
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: PCT/JP91/01696
17 FILING DATE: 06-DEC-1991
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: JP 407922/1990
20 FILING DATE: 27-DEC-1990
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: JP 078840/1991
23 FILING DATE: 11-APR-1991
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: JP 140051/1991
26 FILING DATE: 12-JUN-1991
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: PCT/JP92/00739
29 FILING DATE: 10-JUN-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: JP 212692/1992
32 FILING DATE: 10-AUG-1992
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: PCT/JP93/01101
35 FILING DATE: 05-AUG-1993
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: JP 340078/1992
38 FILING DATE: 21-DEC-1992
39 ATTORNEY/AGENT INFORMATION:
40 NAME: Wegener, Harold C.
41 REGISTRATION NUMBER: 25,258
42 REFERENCE/DOCKET NUMBER: 74129/127/AOPA
43 TELECOMMUNICATION INFORMATION:
44 TELEPHONE: (202) 672-5300
45 TELEFAX: (202) 672-5399
46 TELEX: 904136
47 INFORMATION FOR SEQ ID NO: 5:
48 SEQUENCE CHARACTERISTICS:
49 LENGTH: 1785 base pairs
50 TYPE: nucleic acid
51 STRANDEDNESS: double
52 TOPOLOGY: linear
53 ORIGINAL SOURCE:
54 STRAIN: JM109 PAD404 (FERM BP-3913)
55 FEATURE:
56 NAME/KEY: CDS
57 LOCATION: 233..1144
58 US-08-294-871A-5
59
60 Alignment_scores:
61 Quality: 208.00 Length: 343
62 Ratio: 1.292 Gaps: 14
63 Percent Similarity: 46.939 Percent Identity: 23.907
64
65 Alignment_block:
66 US-09-357-675C-22 x US-08-294-871A-5 ..
67
68 Align seg 1/1 to: US-08-294-871A-5 from: 1 to: 1785
69
70 46 AAlaValCysGlnValThr.....SerThrProAs 55
71 |||:||||| :|||:|||||
72 164 GCGAGCTGTTTCACATCTTTCCGCGCGCTGGCCGACATCACCCTAG 2131

```

```

55 nlysgIngluaSnPhelystrh..... 62
   |||:|||||:|||||:|||||
214 GAAGGAGGAGAGGTTTCATGACACGTCAGATGATTCTTCAGTCGGGACAAAC 263
63 .....Cys 63
264 AAGGTCCGATCGCGCGCGGAGACACGCGAACAGGTGCTGCTGCTT 313
64 ALaGluLeuValGInGluLaLaLaatGLeuGlyAlaCysLeuAlaPheLe 80
   |||:|||||:|||||:|||||:|||||
314 CTCGACATGCTGACGAAGCCCGGACCGCGGCGCAATTCATTCTCT 363
80 uproGlu.....AlaPheAspPheLaLaLa 89
   |||:|||||:|||||:|||||
364 CCCCAGACTCGCGCTTACGACCTTCTCCCGCGCTGCAATTCACCGAGC 413
89 rGAsnProLaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
   |||:|||||:|||||:|||||:|||||
414 AGCGCGAGCTGATGCTTCTATGACAGCGAA...ATGCGCGCGCGCGTG 460
106 LeuGlyGInTyrSerGInLeuAlaArgGlyCysGlyIleTrpLeuSerLe 122
   |||:|||||:|||||:|||||:|||||
461 GTCCGTCACCTCTTGAGAGGCGCGGAACTCGGATCGGCTTCACTCT 510
122 u.....GlyGlyPheHisGluArgGlyGlna 131
   |||:|||||:|||||:|||||
511 GGGCTACGCTGACCTGCTGCGAGAGCGGCTGCAAGCGTCGC..... 553
131 sPTRpGluGInAsnGInLysIleTyrAsnCysHisValLeuLeuAsnSer 147
   |||:|||||:|||||:|||||:|||||
554 .....TTCACACGCTCCATTTTGGTGCGATTAAG 580
148 LysGlySerValValAlaSerTyrArgLysTyrHisLeuCysAspValG 164
   |||:|||||:|||||:|||||:|||||
581 TCAGCGCAATGCTGCGCAAGTATCGTAAGATCCATTTC..... 619
164 uLleProGlyGIn.....GlyProMetArg.....G 173
   |||:|||||:|||||:|||||:|||||
620 ....CCGGGTCAACAGAGTACGAGCGCTACCGCGCTTCACGATCTTG 665
173 LuSerAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
   |||:|||||:|||||:|||||:|||||
666 AAAGCGTATTTCGAGCGCGGATCTCGGCTCCCGGCTGATGACGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
   |||:|||||:|||||:|||||:|||||
716 GACGCGCGAATAATGGGATGTTTCATCTGCAACGATCGCGCTGCGCTGA 765
206 uLeuSerLeuLysLeuAlaGluLaLaGluIleLeuThyTrpProS 223
   |||:|||||:|||||:|||||:|||||
766 ACCCTGCGCGGTATGGCTCAGGGGCGCGGAGATCATCTGC..... 808
223 eAlaPheGlySerValThrGlyProAlaHisTrpGluValLeu..... 237
   |||:|||||:|||||:|||||:|||||
809 .....GGCGGCTACACAGCCGCGACCCACAATCCCTTGTTCGCCAG 850
238 .....LeuArgLaLaArgAl 242
   |||:|||||:|||||:|||||:|||||
851 CACGACACCTGACGTCCTTCCACATCTCTATGATGAGCGCGGCTG 900
242 AlLeuGluSerGInCysTyrValIleAlaLaLaGInCysGlyArgHis 259
   |||:|||||:|||||:|||||:|||||
901 TTATTCAGAGCGGCGCTGCGCGCGCGGAGTGCGCATG...G 947
259 IsGluThrArgAlaSerTyrGlyHisSerMetValValaAspProTrpGly 275
   |||:|||||:|||||:|||||:|||||
948 AGGAGAACTGCATGCTGCTGCGCACTCTCGCATCTGGGCGCGAGCGG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaAr 291
   |||:|||||:|||||:|||||:|||||
998 GAAATGTCGCTCTCACTAGACGCTGAAAGAGAGAGTGATCACCGCGG 1047

```

```

291 gIleAspLeuHisPheLeuGInMeTarGInHisLeuProValPhe. 307
   |||:|||||:|||||:|||||:|||||
1048 CGTCGATCTGATCGCTGCGCGGAGACTGCTGAACACATCTTCACTTCA 1097
308 ..GInHisArgArgProAspLeuTyrGly 316
   |||:|||||:|||||:|||||:|||||
1098 AGCAGCATGCTGACGCCCGCACGACTATGCT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-5
seq_documentation_block:
; Sequence 5, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKABA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25, 258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 pAD404 (FERM BP-3913)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-5

```

alignment_scores:

Quality: 208.00 Length: 343
 Ratio: 1.292 Gaps: 14
 Percent Similarity: 46.939 Percent Identity: 23.907

alignment_block:

US-09-357-675c-22 x US-08-876-398A-5

Align seg 1/1 to: US-08-876-398A-5 from: 1 to: 1785

```

46 AlaValCysGlnValThr.....SerThrProAs 55
|||||:|||||:|||||:|||||:|||||:|||||:
164 GCGAGTGTGTTCCACAACAGCTTTCCCGCGCGCTGGCGCGGACATCAGCTAG 213
|||||:|||||:|||||:|||||:|||||:|||||:
55 nLysGlnGluAsnPhelysThr..... 62
|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACACGTCATGATGACTTGCACTGGGACAAAC 263
|||||:|||||:|||||:|||||:|||||:|||||:
63 .....Cys 63
264 AAGTTCGATCGCGCGCGGAGACAGCGAAGAGTGTGCTGCTT 313
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80
|||||:|||||:|||||:|||||:|||||:|||||:
314 CTGCACTGCTGACGAAAGCCGCGAGCCGCGCAATTTCATGCTCTT 363
|||||:|||||:|||||:|||||:|||||:|||||:
80 uProGlu.....AlaPheAspPheLeAla 89
|||||:|||||:|||||:|||||:|||||:|||||:
364 CCCCAGACCTCGCGCTTACGACCTTCTCCCGCGCTGGCATTTCCACCGAGC 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
|||||:|||||:|||||:|||||:|||||:|||||:
414 AGCGCGAGCTCGATGACTTCTATGAGACCGAA...ATGCCCGCGCGG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGlyCysGlyIleThrLeuSerLe 122
|||||:|||||:|||||:|||||:|||||:|||||:
461 GTCCGCTCCACTCTTTGAGAGGCGCGGAACTCGGATCGGCTTCAATCT 510
122 u.....GlyIlePheHisGluArgGlyGln 131
|||||:|||||:|||||:|||||:|||||:|||||:
511 GGGCTACAGCTGAACTCGTCTGAAAGCGCGCGCTCAAGCGCTCGC..... 553
131 sPTpGluGlnAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 147
|||||:|||||:|||||:|||||:|||||:|||||:
554 .....TTCACACAGTCATTTTGGTGATAG 580
148 LysGlySerValValAlaSerTyrArgGlyThrHisLeuCysAspValAl 164
|||||:|||||:|||||:|||||:|||||:|||||:
581 TCAGGCAAGATCGTCGCGCAAGTATGTAAGATCCATTTG..... 619
164 uLeuProGlyGln.....GlyProMetArg.....G 173
|||||:|||||:|||||:|||||:|||||:|||||:
620 ....CCGGGTCAAGAGAGTACGAGCGCTACCGCGCTTCCAGCATTTG 665
173 LuSerAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
|||||:|||||:|||||:|||||:|||||:|||||:
666 AAAAGCGTATTTGAGCGCGGCGATTCGGGCTCCCGGTCAATAGCGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
|||||:|||||:|||||:|||||:|||||:|||||:
716 GACGCGCGGAAATGGGATGTTCTCCCAACAGATCCCGCGCGCTG 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyValAlaGluLeuThrTyrPro 223
|||||:|||||:|||||:|||||:|||||:|||||:
766 AGCTTCGGCGGATGGCGCTGAGGGCGCGGAGATCATCTGC..... 808
223 eAlaPheGlySerValThrGlyProAlaHisThrGluValLeu..... 237
|||||:|||||:|||||:|||||:|||||:|||||:
809 .....GGCGGTAAACAACAGCGGACCCACAAATCCCTTTTCCCGG 850
238 .....LeuArgAlaArgAl 242
|||||:|||||:|||||:|||||:|||||:|||||:

```

```

851 CACGACCACCTGACGTCCTCCACCATCTCTATGATGACGCGCGGTC 900
242 atLeuSerGlnCysTyrValIleAlaAlaIleGlnCysGlyArgHis 259
|||||:|||||:|||||:|||||:|||||:|||||:
901 TTATCAGAAAGGCGCTGCTCCGCGCGCGCGGCAAGTGGGATG...G 947
259 isGluThrArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 275
|||||:|||||:|||||:|||||:|||||:|||||:
948 AGGAGAACTGCATGCTGCGCGGCACTCTGCATCTGGCGCGGACCGG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
|||||:|||||:|||||:|||||:|||||:|||||:
998 GAATCGTGCTCTCACTACGACGCTGGAAGACGAGTATCACCAGCGG 1047
291 gileAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307
|||||:|||||:|||||:|||||:|||||:|||||:
1048 CGTGCATCTGATCGCTCGCGGAACTGGGTGTAACATCTTCAACTTCA 1097
308 ..GlnHisArgArgProAspLeuTyrGly 316
|||||:|||||:|||||:|||||:|||||:|||||:
1098 AGCAGCATGTCACCCGACCACTATGAT 1126
seq_name: /cgn2_6/ptodate/2/ina/5A_COMB.seq:US-08-294-871A-61
seq_documentation_block:
Sequence 61, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanda, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshio
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990

```

```

414 AGGCCGAGCTGCATGACTTCTGTATGAGACCGAA...ATGCCGCGGCCCGGTG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGlyCysGlyIleTrrPheSerLe 122
      :::: ||| ||| ||||| :::: |||
461 GTCCGCTGCACCTCTTTGAGAAAGCCGCGGAAATCGGGATCGGCTTCAATCT 510
122 u.....GlyGlyPheHisGlnAArgGlyGlnA 131
      | ||||| ||||| :::: |||
511 GGGCTACGCTGAATCTGCTGCGAAGGCGCGCTACGCTGC..... 553
131 sPTrrGlnGlnAsnGlnIleTyrAsnGlyHisIleLeuAsnSer 147
      :::: ||| :::: ||| :::: ||| :::: |||
554 .....TTCAACAGCTGCATTTTGGTGGATPAG 580
148 LysGlySerValAlaAlaSerTyrArgGlyThrHisLeuCysAspValG1 164
      :::: ||| :::: ||| :::: ||| :::: |||
581 TCAGCGAAGATCTGTCGGCAGATGTCGTAAGATCCATTGG..... 619
164 uIleProGlyGln.....GlyProMetArg.....G 173
      | ||||| :::: ||| :::: |||
620 ...CCGGCTCAACAGAGATGACAGGCTTACCGGCGCTTCAGAGATCTTG 665
173 IuserAsnTyrThrLysProGlyGlyThrLeuGlnPrrProValLysThr 189
      ||:||||| :::: ||| :::: |||
666 AAAAGCGTTATTTGCAAGCCGGGAGATCTCGGCTTCCGGCTTAAGAGCTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG1 206
      |||:||||| :::: ||| ||| :::: |||
716 GACGCGCGCGAAATGGGGATGTCATCTGCACAGATCGCGCTGAGCTGA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyAluIleLeuThrTyrProS 223
      | :::: ||| ||||| ||||| :::: |||
766 AGCGTCGGCGGTATGGGCTCAGGGGCGCGAGATCATCTGC..... 808
223 eAlaLarPheGlySerValThrGlyProAlaHisTrrGlnValLeu..... 237
809 .....GGCGCTACACAGCGCGCACCAATCCCGCTGTTCCCGAG 850
238 .....LeuArgAlaArgAl 242
851 CACGACCACTGACAGTCTTCCACCAATCTCTATGATGACAGCGCGGTG 900
242 aIleGlySerGlnCysTyrValIleAlaAlaAlaGlnCysGlyArgHisH 259
      : :::: ||| :::: ||| :::: |||
901 TTATCGAAGACGGGCGCTGTGTCCGGCGCGCGGCGCAAGGTGGGCAATG..G 947
259 IsgIuThrArgAlaSerTyrGlnHisSerMetValAlaAspProTrrGly 275
948 AGGAGAACTGCATGCTGCTGGCGCACTCTGCATGTGTGGCGCGACCGGG 997
276 ThrValValAlaArgCysSerGly...GlyProGlyLeuCysLeuAlaAr 291
      :::: ||| :::: ||| :::: |||
998 GAATGCTGCTCTCACTACAGACCTTGAAGAGACAGATGATACACCGCGC 1044
291 gIleAsrPheLnisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307
      :::: ||| :::: ||| :::: ||| :::: |||
1048 CCGTGAATCTGCATGCTGCGCGGAACTGCGTGAACACATCTTCAACTTCA 1097
308 ..GlnHisArgArgProAsrPheuTrrGly 316
      | ||||| ||||| :::: |||
1098 ACCAGCATCTTCAGCCCGCACACTACTGT 1126
seq_name: /cgn2_6/plodata/2/lna/6A.COMB.seq:US-08-876-398A-61
seq_documentation_block:
; Sequence 61, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio

```

APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD468
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-61

alignment_scores:
Quality: 207.00 Length: 343
Ratio: 1.286 Gaps: 14
Percent Similarity: 46.939 Percent Identity: 23.907

alignment_block:
US-09-357-675C-22 x US-08-876-398A-61 ..

Align seg 1/1 to: US-08-876-398A-61 from: 1 to: 1785

46 AAlaValCysGlnValThr.....SerThrProAs 55
|||||:|||||:|||||:|||||:|||||:|||||:
164 GCGATGTGTTTCAACAGCTTTCCGCGCGCTGGCGGACATCACCTAG 213
55 nlysgIngluAsnPhelysThr..... 62
:|||||:|||||:|||||:|||||:|||||:|||||:
214 GAAGGAGCAGAGGTTTCATGACAGTCAGATGATCTTGACGTGGGACAC 263

63Cys 63
264 AAGCTCGATCGCGCGCGAGACACGCGAAGAGGTGCTGCTCTT 313
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
:|||||:|||||:|||||:|||||:|||||:|||||:
314 CTCGACATGCTGACGAAGCCGCGACCGGGCGCAATTTCAATTCCTT 363
80 uProGlu.....AlaPheAspPheLeuAla 89
:|||||:|||||:|||||:|||||:|||||:|||||:
364 CCGCGAATCGCGCTTACGACCTTTCCCGCGCTGACATTTCACGAG 413
89 rGAsnProAlaGluThrLeuLeuSerGluProLeuAsnGlyAspLeu 105
:|||||:|||||:|||||:|||||:|||||:|||||:
414 AGCCGAGCTGATGATCTTATGAGACCGAA...ATGCCCGCGCGGTG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGluCysGlyIleThrLeuSer 122
:|||||:|||||:|||||:|||||:|||||:|||||:
461 GTCCGTCACCTTTTGAGAGCGCGGGAATCGGATCGGCTCAATCT 510
122 u.....GlyGlyPheHisGluArgGlyAla 131
:|||||:|||||:|||||:|||||:|||||:|||||:
511 GGGCTACGCTGAACCTGCTGCGAAGCGCGGTCAAGCTGCG..... 553
131 sPrpGluGlnAsnGlnLysIleTyrAsnCysHisValLeuAsnSer 147
:|||||:|||||:|||||:|||||:|||||:|||||:
554TTCACACGTCATTTTGGTGATAG 580
148 LysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspValG 164
:|||||:|||||:|||||:|||||:|||||:|||||:
581 TCAGGCAGATGCTGCGCAATCTGTAAGATCCATTG..... 619
164 uIleProGlyGln.....GlyProMetArg.....G 173
:|||||:|||||:|||||:|||||:|||||:|||||:
620 ...CCGGGTCAAGAGATGACGAGCTTACCGCGCTTCCAGCATCTTG 665
173 LysAsnTyrThrLysProGlyGlyThrLeuLupProValLysThr 189
:|||||:|||||:|||||:|||||:|||||:|||||:
666 AAAGCGTTATTTCGAGCGCGGATCTGCGGTCCGCTATGACGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPhePro 206
:|||||:|||||:|||||:|||||:|||||:|||||:
716 GAGCCGCGCAAAATGGGATGTCATCTGCACGATCGCGCGCTGCTGA 765
206 uLeuSerLeuLysLeuAlaGluAlaGlyAlaGluIleLeuThrTyros 223
:|||||:|||||:|||||:|||||:|||||:|||||:
766 AGCTGCGCGGTGATGGCTCAGGGGCGCGGAGATCATCTGC..... 808
223 eAlaPheGlySerValThrGlyProAlaHisTPrpLysValLeu..... 237
:|||||:|||||:|||||:|||||:|||||:|||||:
809GGCGGTACAAACGCGCAACGCAATCGCGCTGTCCCGAG 850
238LeuArgAlaArgAla 242
:|||||:|||||:|||||:|||||:|||||:|||||:
851 CACGACACCTGACGTCCTTCCACATCTCTATGCGTACAGCGCGGTC 900
242 aIleGluSerGlnCysTyrValIleAlaAlaIleGlnCysGlyArgHis 259
:|||||:|||||:|||||:|||||:|||||:|||||:
901 TTATTCGAACGGGGCTGTGTCGCGCGCGCGGCAAGGTGGGATG... 947
259 IsGluThrArgAlaSerTyrGlyHisSerMetValValAspProTPrp 275
:|||||:|||||:|||||:|||||:|||||:|||||:
948 AGGAGAACTGCATGCTGTCGCGCTGCTGCGCGCGGAGGAGG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
:|||||:|||||:|||||:|||||:|||||:|||||:
998 GAATGTCGCTCTCACTACGAGCTGGAAGAGAGAGATATCACCGCCG 1047
291 gIleAspLeuHisPheLeuGlnIleMetArgGlnHisLeuProValPhe 307
:|||||:|||||:|||||:|||||:|||||:|||||:
1048 CGTGATCTCGATCGCTCGCGGAACTGCGCTGAAACATCTTCACTTGA 1097
308 ..GlnHisArgArgProAspLeuTyrGly 316

```

|||||
1098 AGCAGCATGTCAGCCCCAGCACTATAGCT 1126

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-294-871A-7
seq_documentation_block:
Sequence 7, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanda, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
TITLE OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992

```

```

ATTORNEY/AGENT INFORMATION:
1 NAME: Wegner, Harold C.
2 REGISTRATION NUMBER: 25,258
3 REFERENCE/DOCKET NUMBER: 74129/127/AOPA
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: (202) 672-5300
6 TELEFAX: (202) 672-5399
7 TELEX: 904136
8 INFORMATION FOR SEQ ID NO: 7:
9 SEQUENCE CHARACTERISTICS:
10 LENGTH: 1785 base pairs
11 TYPE: nucleic acid
12 STRANDEDNESS: double
13 TOPOLOGY: linear
14 ORIGINAL SOURCE:
15 STRAIN: JM109 PAD406 (FERM BP-3914)
16 FEATURE:
17 NAME/KEY: CDS
18 LOCATION: 233..1144
19 US-08-294-871A-7

alignment_scores:
20 Quality: 206.00 Length: 338
21 Ratio: 1.280 Gaps: 14
22 Percent Similarity: 47.633 Percent Identity: 24.260

alignment_block:
23 US-09-357-675C-22 x US-08-294-871A-7 ..

Align seg 1/1 to: US-08-294-871A-7 from: 1 to: 1785

46 AAlaValcSgInValThr.....SerThrProAs 55
47 |||:||||| :|||:|
48 164 GCATGTGTTTCAACAAGGTTTCCCGCGCGTGGCGCGACATCACTAG 213
49
50 55 nlySgInGlnAsnPhelyThr..... 62
51 :|||:|:|:|:|:| |
52 214 GAAGGAGACAGAGGTTTCATGACACGTCAAGTATGATCTTGCATGGGACAA 263
53
54 63 .....Cys 63
55
56 264 AAGGTCCGATCGCGCGCGGAGACACCGGAACAGTGTCTGTCTT 313
57
58 64 AAlaGluValGlnGlnAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
59 :|||:|:|:|:|:| |
60 314 CTGACACTGCTGACGAAGCCGCCGAGCGCGCGGATTTATTGTCTT 363
61
62 80 uProGlu.....AlaPheAspPheIleAla 89
63 :|||:|:|:|:| |
64 364 CCCCAGAACTGGCGGTACAGACTTCTTCCGCGCGTGGCATTTACACGAG 413
65
66 89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAsp 105
67 :|||:|:|:|:|:| |
68 414 AGCCCGAGCTCGAAGAGCTTCATGAGACGAA..ATGCCCGCCCGGTG 460
69
70 106 LeuGlyAlaThrSerGlnLeuAlaArgCysGlyIleThrPheSer 122
71 :|||:|:|:|:|:| |
72 461 GTCCGTCACACTTTTGAGAGAGCCGCCGGAACCTCGGGATGGCTTCAAT 510
73
74 122 u.....GlyIlePheAsnGluArgGlyGln 131
75 :|||:|:|:|:| |
76 511 GGGCTACGTGAACTCGTCTGGAAGCGCGGTCAAGCTGC..... 553
77
78 131 sprrrrGlnGlnAsnGlnIleThrGlyAsnCysHisValLeuLeuAsn 147
79 :|||:|:|:|:|:| |
80 554 .....TTCAACAGCTGCATTTTGGTGGATAAG 580
81
82 148 LysGlySerValAlaIleSerThrArgGlyThrHisLeuAspArgAla 164
83 :|||:|:|:|:|:|:| |
84 581 TCAAGCAAGATCGTGGCAAGTATCGTAAGATCCATTG..... 619
85
86 164 uileProGlyGln.....GlyProMetArg.....G 173

```

```

620 .....|
173 luserstnryrthlysproglygylthrlengulprprovalylsthr 169
666 AAAAGCGTATTTCAGACCGCGGATCTCGCTCCGCTCTATGACG 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
716 GACGCGCGGAAATGGGATGTTTCATCTGCACAGATCGCGCGCTG 765
206 uleuSerLeuLysLeuAlaIleGlyAlaIleLeuThr.....T 221
766 AGCGTGGCGGATGAGCGCTCAGGCGCGGAGATCATCTCGCGCGCT 815
221 yProSerAlaPheGlySerValThrglyProAlaHstirpGluVal 237
816 ACAACAGCGCGACCCACATCCCTCTGTCGCCAGCAGCACCTGACG 865
238 .....LeuArgAlaArgAlaIleGluSerGlnCy 247
866 TCCTTCACCATCTCTATTCATGACGAGCGCGGCTTATCAGAACGG 915
247 stYrValIleAlaAlaIleGlnCysGlyArgHstirpGluThrArg 264
916 CTGCTCCGCGCGCGGCGGCAAGGTGGGATG...GAGGAGACTGCTG 962
264 ertYrGlyHstirpMetValValAspProtrpGlyThrValAlaI 280
963 TGCTCGGCGCTCTCGATCTGCGCGCGGCAAGGTGGGCTGCTC 1012
281 CysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeu 296
1013 ACACAGAGCTGAGACGAGGTGATCACCAGCGCGCTCATCTCATG 1062
296 eleuGlnGlnMetArgGlnHstirpProValPhe...GlnHstirp 312
1063 CTGCGCGGAGACTCGTGACACATCTTCACACTTCAGACACATCG 1112
312 roAspLeuArgly 316
1113 CCCAGCATATGTGT 1126

seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:US-08-294-871A-9

seq documentation block:
; Sequence 9, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nambu, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraiishi, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; NUMBER OF INVENTIONS: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; STRAIN: JM109 PAD416 (FERM BP-3915)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 233..1144
; US-08-294-871A-9

alignment_scores:
Quality: 206.00 Length: 343
Ratio: 1.280 Gaps: 14
Percent Similarity: 46.939 Percent Identity: 23.907

alignment block:
US-09-357-675c-22 x US-08-294-871A-9 ..
Align seg 1/1 to: US-08-294-871A-9 from: 1 to: 1785
46 AlaValCysGlnValThr.....SerThrProAs 55
```


FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/127/AOPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
ORIGINAL SOURCE:
STRAIN: JMI09 PAD428
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1144
US-08-294-871A-11

alignment_scores:
Quality: 206.00 Length: 343
Ratio: 1.280 Gaps: 14
Percent Similarity: 46.939 Percent Identity: 23.907

alignment_block:
US-09-357-675c-22 x US-08-294-871A-11 ..

Align seg 1/1 to: US-08-294-871A-11 from: 1 to: 1785

```
46 AlaValCysGlnValThr.....SerThrProAs 55
|||||.....
164 GCATGTGTTTCACAAACGTTTCCCGCGCGCTGGCCGCGGACATCACCCTAG 213
55 nLysGlnIAsnPhelYThr..... 62
|||||.....
214 GAAGAGACAGAGGTTTCATGACAGTCAGATGACTTGCAGTGGGACAC 263
63 .....Cys 63
264 AAGTCGATCGCGCGCGGAGACACGCGACAGTCTGTCGTTCTT 313
64 AlaGlnValGlnGlnAlaIaIaArgLeuGlyAlaCysLeuAlaPhe 80
|||||.....
314 CTCGACATGTCGACGAAGCGCGGACCGGGCGCGAATTCTTCTT 363
80 uProGlu.....AlaPheAspPheIleAla 89
|||||.....
364 CCGCGAAGTCGGCGCTTACGACCTTCTCCCGCGCTGGCATTACCGAGC 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
|||||.....
414 AGCCCGAGCTCGATACCTTCTATGAGACCGAA...ATGCCCGCGCGGTG 460
105 LeuGlyGlnTyrSerGlnLeuAlaArgIuGlyGlyTyrLeuSerLe 122
|||||.....
461 GTCCTGCACCTTTTGAAGGCGCGGAACTCGGATCGGCTTCAATC 510
122 u.....GlyGlyPheHisGlyArgGlyGln 131
|||||.....
511 GGGCTACGCTGAACCTGCTGCAAGCGCGCGTCAAGCGTCC..... 553
131 sPTpGluGlnAsnGlnIleTyrAsnCysHisValLeuLeuAsnSer 147
|||||.....
```

```
554 .....TTCACACGTCATTTTGGTGATAC 560
148 LysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspAla 164
|||||.....
581 TCAGCGAAGATCGTCGCAAGTATCTGAATCATTTG..... 619
164 uileProGlyGln.....GlyPrometArg.....G 173
|||||.....
620 ....CCGGGTACACAGAGATACGAGGCTTACCGCGCTTCCACACATTTG 665
173 LysSerAsnTyrThrLysProGlyGlyThrLeuGluProValLysThr 189
|||||.....
666 AAAAGGTTATTTTCAGCCGCGGATCTCGGCTCCGCTTATGACATC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPhePro 206
|||||.....
716 GACCGCGCGAAATGGGATGTTTCATCTGACAGATCGCGCGCTGTA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGluIleLeuThrTyrPro 223
|||||.....
766 AGCCTGCGGCTGATGGCTCAGGGCGCGGAGATCATCTGC..... 808
223 eAlaPheGlySerValThrGlyProAla..... 232
|||||.....
809 .....GGCGGTACACACGCGGACCCACATCCCTGTTCCCGAG 850
233 .....HisTrpGluValLeuLeuArgAla 242
|||||.....
851 CACGACCACCTGACGTCCTCCACCATCTTTCATGACGAGCGCGGTC 900
242 aileGlySerGlnCysTyrValIleAlaIaIaGlnCysGlyArgHis 259
|||||.....
901 TATTCAGAACGGGGCTGTCGCGCGCGCGGCGGAGGCGCGCATG...G 947
259 isGluThrArgAlaSerTyrGlyHisSerMetValAlaAspProTrpGly 275
|||||.....
948 AGGAGAACTGCATGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 997
276 ThrValValAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
|||||.....
998 GAATGTCGCTCTCCTACAGCGCTGGAAGACGAGTATCATCCGCC 1047
291 gileAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307
|||||.....
1048 CGTCGATCTCGATCGCTCGCGGAACTGCGGTGAACATCTTCACTTCA 1097
308 ..GlnHisArgArgProAspLeuTyrGly 316
|||||.....
1098 AGCAGCATGCTCAGCGCCAGACACTATGT 1126
```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-294-871A-21

seq_documentation_block:
Sequence 21, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Namba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiraiishi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.

```

? COUNTRY: U.S.
? ZIP: 20007-5109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/294,871A
? FILING DATE: 22-AUG-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/971,758
? FILING DATE: 12-APR-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/917,111
? FILING DATE: 07-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/211,641
? FILING DATE: 11-APR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 400848/1990
? FILING DATE: 07-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP91/01696
? FILING DATE: 06-DEC-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 407922/1990
? FILING DATE: 27-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 078840/1991
? FILING DATE: 11-APR-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 140051/1991
? FILING DATE: 12-JUN-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP92/00739
? FILING DATE: 10-JUN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 212692/1992
? FILING DATE: 10-AUG-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/JP93/01101
? FILING DATE: 05-AUG-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 340078/1992
? FILING DATE: 21-DEC-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Wegner, Harold C.
? REGISTRATION NUMBER: 25,258
? REFERENCE/DOCKET NUMBER: 74129/127/AOPA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 672-5300
? TELEFAX: (202) 672-5399
? TELEEX: 904136
? INFORMATION FOR SEQ ID NO: 21:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1785 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? ORIGINAL SOURCE:
? STRAIN: JM109 PAD439
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 233..1144
? US-08-294-871A-21

```

```

alignment_scores:
  Quality: 206.00      Length: 343
  Ratio: 1.280        Gaps: 14
  Percent Similarity: 46.939  Percent Identity: 23.907

```

```

alignment_block:
US-09-357-675c-22 x US-08-294-871A-21 ..
Align seg 1/1 to: US-08-294-871A-21 from: 1 to: 1785

46 AlaValCysGlnValThr.....SerThrProAs 55
164 GCGATGTGTTTCAACACGTTTCCCGGCGCTGGGCGGACATCACTAG 213
55 nlySGlnGluAsnPhelyThr..... 62
214 GAAAGACGACAGAGTTTCATGACACGTCAGATGATTAATTGACGTGGACAAC 263
63 .....Cys 63
264 AAGTCCGATCGCGCGCGGAGACGCGAACAGTCGTCTGCTCTT 313
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
314 CTCGACATGCTGACGAAAGCGCGAGCGCGCGCATTTTCATTGCTT 363
80 uProGlu.....AlaPheAspPheIleAla 89
364 CCCGAACTCGCGCTTACGACCTTCTCCGCGCTGCGATTTCACGACG 413
89 rGAsnProAlaGluThrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
414 AGCGCGAGCTCGATAGCTTCTATGACCGAA...ATGCCGCGCGGTG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGluCysGlyIleTyrLeuSer 122
461 GTCCGTCACACTTGTGAAAGAGCGCGGACGTCGGATTCGCTTCAATCT 510
122 u.....GlyGlyPheHisGluArgGlyGln 131
511 GGGCTACGCTGAACTCGTCGTCGAAGCGCGCTCAAGGTCG..... 553
131 sPrTprGluGlnAsnGlnLysIleTyrAsnGlyHisValLeuLeuAsnSer 147
554 .....TTCACACGCTCATTTTGTGATTAAG 580
148 LysGlySerValAlaAlaSerTyrArgLysThrHisLeuCysAspValG 164
581 TCAGGCAAGATCGTCGGAATATCGTAAGATTCATTG..... 619
164 uIleProGlyGln.....GlyProMetArg.....G 173
620 ....CCGGTCCACAAGAGTACGAGGCTTACCGCGCTTCCAGCATCTTG 665
173 LuserAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
666 AAAAGCGTTATTTCGAGCCGGCGATCTCGGCTTCCCGGTATAGACGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
716 GAGCCCGGAAATGGGATGTCATGCAACGATCCCGCGCTCGA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 223
766 AGCCTGGCGGGTATGGCTCAGAGGGCGCGCATCATCTGC..... 808
223 eAlaPheGlySerValThrGlyProAla..... 232
809 .....GGCGGCTACACAGCGCGACCCACATCCCGCTTCCCGAG 850
233 .....HisTrpGluValLeuLeuArgAlaArgAl 242
851 CAGACCACTGACGTCCTTCCACCATCTCTATGATGACGCGCGGTC 900
242 aIleGluSerGlnCysTyrValIleAlaAlaGlnCysLysArgHisH 259
901 TTATCAGAAAGCGGCTGTCGCGCGCGCGCAAGCTGTCATG...G 947

```

```

259 1sgluThrArgAlaSerTYGLYHisSerMetValAlaSprTrpGly 275
      ::::: ||||| :::: |||
948 AGAGAGACGATCGTCGGCGACTCCGCTGCGCGCGGACGGG 997
      ::::: ||||| :::: |||
276 ThrValAlaAlaArgCysSerGlu...GlyProGlyLeuCysLeuAla 291
      ::::: ||||| :::: |||
998 GAATGTGCTCTCTCACTAGCAGCTGTAAGACGAGCTGTCACCCCGC 1047
      ::::: ||||| :::: |||
291 gltAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe 307
      ::::: ||||| :::: |||
1048 CGTCGATCTCGATCGCTCGCGGAGACTGCTGAACACATCTTCACTCA 1097
      ::::: ||||| :::: |||
308 ..GlnHisArgArgProAspLeuTYrGly 316
      ::::: ||||| :::: |||
1098 AGCAGCATGTCGACGCCCAAGCACTATGCT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-59

seq_documentation_block:
: Sequence 59, Application US/08294871A
: Patent No. 5824522
:
: GENERAL INFORMATION:
: APPLICANT: Ikemaka, Yasuhiro
: APPLICANT: Nanda, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
: APPLICANT: Yamada, Kazuhiko
: APPLICANT: Hiratschi, Yoshiro
: TITLE OF INVENTION: Immobilized Enzyme Preparation and
: TITLE OF INVENTION: Process for Producing D-a-Amino Acid
: NUMBER OF SEQUENCES: 70
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.
: ZIP: 20007-5109
:
: COMPUTER-READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/294,871A
: FILING DATE: 22-AUG-1994
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,758
: FILING DATE: 12-APR-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/917,111
: FILING DATE: 07-AUG-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/211,641
: FILING DATE: 11-APR-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 400848/1990
: FILING DATE: 07-DEC-1990
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP91/01696
: FILING DATE: 06-DEC-1991
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 407922/1990
: FILING DATE: 27-DEC-1990
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 078840/1991
: FILING DATE: 11-APR-1991
:
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: JP 140051/1991
: FILING DATE: 12-JUN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP92/00739
: FILING DATE: 10-JUN-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 212692/1992
: FILING DATE: 10-AUG-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP93/01101
: FILING DATE: 05-AUG-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 340078/1992
: FILING DATE: 21-DEC-1992
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Wegner, Harold C.
: REGISTRATION NUMBER: 25,258
: REFERENCE/DOCKET NUMBER: 74129/127/AOPA
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 672-5300
: TELEFAX: (202) 672-5399
: TELEX: 904136
:
: INFORMATION FOR SEQ ID NO: 59:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1785 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: ORIGINAL SOURCE:
: STRAIN: JM109 PAD456
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 233..1144
:
: US-08-294-871A-59

```

```

alignment_scores:
      Quality: 206.00      Length: 342
      Ratio: 1.280        Gaps: 15
      Percent Similarity: 47.076      Percent Identity: 24.854

```

alignment_block:

US-09-357-675c-22 x US-08-294-871A-59 ..

Align seg 1/1 to: US-08-294-871A-59 from: 1 to: 1785

```

46 AlaValCysGlnValThr.....SerThrProAs 55
      ||::||| ::|||
164 GCGATGTGTTTCACACAGTTTCCCGCGCTGGCGGAGACATCACCTAG 213
      ::||::|||::|||
55 nlysgIngluAsnPhelysThr..... 62
      ::||::|||::|||
214 GAGGAGCAGAGGTTTCATGACACGTCAGATGATCTTCAGTGGGACAC 263
      ::||::|||::|||
63 .....Cys 63
264 AAGGTCGATTCGGCGCGGAGACACAGGAGAGTCTGCTCTT 313
      ::||::|||::|||
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 80
      ::||::|||::|||
314 CTCGACATCTGTCAGCAAGCCGCGAGCCGGGGCGAATTTCATTGCTTT 363
      ::||::|||::|||
80 uproGlu.....AlaPheAspPheIleAlaArg.....Asp 91
      ::||::|||::|||
364 CCGCGAATCGCGCTTACGACCTTCTTCCCGCGCTGTGTTTTCACCGAG 413
      ::||::|||::|||
91 roAlaGluThrLeuLeuLeuSerGlu...ProLeuAsnGlyAspLeu 106
      ::||::|||::|||
414 AGGCGAGCTCGATAGCTTCTTATAGAGACGGAATGCGCGCGGATGTC 463
      ::||::|||::|||
107 GlyGluTYrSerGlnLeuAlaArgGluCysGlyIleTrpLeuSerLeu.. 122
      ::||::|||::|||
464 CTCGACCTTTGAGAAAGCCCGGAACATTCGGGATCGCTTCAATCTGGG 513

```

```

123 ..... GlyGlyPheHisGluArgGlyGlnAspT 132
514 CTACGCTGAACCTGCTCGTCGAGAGCGCGTCAAGCGCTGC..... 553
132 rpluGlnAsnGlnLysIleTyrAsnGlyHisValLeuLeuAsnSerLys 148
554 ..... TTCACACGTCATTTTGGTGATTAAGTCA 563
149 GlySerValAlaIAserTyrArgLysThrHisLeuGlyAspValGluI 165
584 GGCAGAGATCGTCGCGCAAGTATCGTAAGATTCATTG..... 619
165 eProGlyGln..... GlyPrometArg..... GluS 174
620 ..CGGGTCACAGAGAGTACGAGCGCTACCGCGCTCCACACTCTTGAA 668
174 eTAsnTyrThrLysProGlyGlyThrLeuGluProValLysThrPro 190
669 AGCGTATTTCGAGCGCGGATCTCGGCTTCCGGCTTATGACGTGAC 718
191 AlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProGlu 207
719 GCCGCGAAATGGGATGTTTCATCTGCAAGCATCGCGCTGCGCTGAAGC 768
207 uSerLeuLysLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProsera 224
769 CTGGCGGGTATGGGCTCAGAGCGCGGAGATCATCTGC..... 808
224 IapheGlySerValThrGlyProAlaHisTrpGluValLeu..... 237
809 ..... GCGGCTTACAAACACGCCGCCACCAATCCCTGTTCCCGCAC 853
238 ..... LeuArgAlaArgAlaI 243
854 GACCACCTGACGTCCTTCACACATCTCTATGATGACAGCGCGGCTCTA 903
243 eGluSerGlnCysTyrValIleAlaAlaGlnGlyArgHisHisG 260
904 TCAGAACGGGGCTGTCCGCGCGGCGCAAGTCAGGCATG...GAGG 950
260 IuThrArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 276
951 AGAATCGCATGCTGCTCGGCCACTCCCTGCAATCGTGGCGCGGAGGAA 1000
277 ValValAlaArgCysSerGlu...GlyProGlyLeuGlySerLeuAlaArgI 292
1001 ATCGTGCCTCTCACTACGACGCTGGAAGACGAGGTATCACCGCGCGCT 1050
292 eAspLeuHisPheLeuGlnGlnMetArgGlnHisLeuProValPhe...G 308
1051 CGATCTCGATGCTGCCGGAGACTGCGTGAACATCTTCAACTTCAAGC 1100
308 IuHisArgArgProAspLeuTyrGly 316
1101 AGCATCGTCAGCCCGACACTATGCT 1126
seq_name: /cgn2.6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-7
seq_documentation_block:
; Sequence 7, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Megner, Harold C.
; REGISTRATION NUMBER: 25, 258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD406 (FERM BP-3914)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-7
alignment_scores:
Quality: 206.00 Length: 338
Ratio: 1.280 Gaps: 14
Percent Similarity: 47.633 Percent Identity: 24.260
alignment_block:
US-09-357-675C-22 x US-08-876-398A-7 ..
Align seg 1/1 to: US-08-876-398A-7 from: 1 to: 1785
46 AlaValCysGlnValThr.....SerThrProAs 55
||||:||||:||||:||||:||||:||||:||||:
164 GCGATGTGTTTCAACACGTTTTCGCGCGCTGGCGGAGCATATCCTAG 213
55 nLysGlnGluAsnPheLysThr..... 62
||||:||||:||||:||||:||||:||||:
214 GAAGAGACAGAGTTTCATGACAGCTCAGATGATCTTGCAGTGGAGAAC 263
63 .....Cys 63
264 AAGTCCGATCGCGCGGAGACAGCGAAGTGTCTGCTTCTT 313
64 AlaGluLeuValGlnGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 80

```

314 CTCGACATCTGACGAAAGCCGAGCCGCGGCGAATTTATGTCCT 363
80 UPFROG.....AlpheasppheilaA 89
364 CCCCAGACTCGCGCTTACGACTTCTCCCGCGCTGGCATTTACCGAG 413
89 rGAspProAlaIuHrLeuLeuLeuSerGluProLeuAsnGlyAspLeu 105
414 AGCCGAGCTCGATAGCTTCTATGAGACCGAA...ATGCCGCGCCGGG 460
106 LeuGlyGlnTyrSerGlnLeuAlaArgGlyGlyIleTyrLeuSer 122
461 GTCGCTGACCTTTGAGAAAGCCGCGGAGACTGGGATCGGCTTCATCT 510
122 U.....GlyGlyPheHisGluArgGlyGlnA 131
511 GGGCTACGCTGAACCTGCTGCAAGCGCGCTCAAGCTGCG..... 553
131 SPTRpGluInAsnGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 147
554TTCAACAGCTCATTTTGGTGATAG 580
148 LysGlySerValValAlaSerTyrArgLysThrHisLeuCysAspValG 164
581 TCAGGCAAGATCGTCGCAAGTATCTAGATCCATTG..... 619
164 uLlePrgLysGln.....GlyPrometArg.....G 173
620 ...CGGGTCACAGAGACTACGAGCCCTACCGCGCTTCACGACTCTTG 665
173 LuserAsnTyrThrLysProGlyGlyThrLeuGluProProValLysThr 189
666 AAAAGCGTATTTCAGCGCGGGGATCTCGGCTCCCGGTCATAGACGTC 715
190 ProAlaGlyLysValGlyLeuAlaIleCysTyrAspMetArgPheProG 206
716 GAGCGCGCAAAATGGGATGTTTCATCTGCACGATCGCGCGGCTGTA 765
206 uLeuSerLeuLysLeuAlaGlnAlaGlnIleLeuThr.....T 221
766 AGCCTGCGCGGGATGGGCTCAGGGGCGCCGAGATCATCTGCGCGGCT 815
221 yProSerAlaPheGlySerValThrGlyProAlaHisTrpGluValLeu 237
816 ACAACAGCGCGAGCCACAAATCCCTGTTCCCGACGACGACGACGAGC 865
238LeuArgAlaArgAlaIleGluSerGlnC 247
866 TCCTTCACACATCTCTATCGATGACGCGCGGCTTATCAGAAAGCGGC 915
247 sTyrValIleAlaAlaGlnCysGlyArgHisHisGluThrArgAla 264
916 CTGGTCGCGCGCGCGCGGCAAGGTGGCATG...GAGGAGAACTGCATGC 962
264 eTyrGlyHisSerMetValValAspProTyrPglyThrValValAlaArg 280
963 TCGTCGCGCACTCTGATCGTCGCGCGGCAAGGAAATCGTCGTCGTC 1012
281 CysSerGln...GlyProGlyLeuCysLeuAlaArgIleAspLeuHisPh 296
1013 ACTACGACGCTGAGAACGACGAGGATCACCGCGCGCTGATCTCGATGC 1062
296 eLeuGlnGlnMetArgGlnHisLeuProValPhe...GlnHisArgArgP 312
1063 CTGCCGGAACATGCGTGAACACATCTTCAACTTCAAGCAGCATGTCAGC 1112
312 roAspLeuTyrGly 316
1113 CCCAGCATATGTT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-9

seq_documentation_block:
Sequence 9, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiro
APPLICANT: NAMBA, Hirokazu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 pad416 (FERM BP-3915)
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-9

alignment_scores:
Quality: 206.00 Length: 343
Ratio: 1.280 Gaps: 14
Percent similarity: 46.939 Percent identity: 23.907

alignment_block:

US-09-357-675c-22 x US-08-876-398A-9 ..

Align seg 1/1 to: US-08-876-398A-9 from: 1 to: 1785

OM of: US-09-357-675c-25 to: GenEmbl:* out_format : pfs
Date: Apr 29, 2002 11:15 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlh
-Q/cg21_1/USPTO.spool/US09357675/rnat_29042002_091757_3179/app.query.fasta.1.1208
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsums2
-TRANS=human40.cdi -LIST=45 -DCCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09357675.ecg1.1.10065 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPRX -WAIT -THREADS=1

Search information block:

Query: US-09-357-675c-25
Query length: 362

Database: GenEmbl:*

Database sequences: 1797656
Database length: 1873333701

Search time (sec): 33/4.980000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pr:AF069987	+ 1336.00	2332.76	1.4e-121	1385	AF069987 Homo sapiens nitrilas
gb_hcg:AL354714-4	+ 1538.50	1915.16	2.6e-98	110000	Continuation (5 of 7) of AL3
gb_hcg:AL354714-5	+ 1538.50	1915.16	2.6e-98	110000	Continuation (6 of 7) of AL3
gb_hcg:AL590651	- 1538.50	1911.42	4.2e-98	167863	AL590651 Homo sapiens chromo
gb_pr:AL591806	+ 1538.50	1909.84	5.1e-98	200822	AL591806 Human DNA sequence
gb_pr:AF069984	+ 1525.50	1927.70	5.2e-99	4079	AF069984 Homo sapiens nitrilas
gb_ro:AF069988	+ 1496.00	1899.97	1.8e-97	1338	AF069988 Mus musculus nitrilas
gb_ro:BC021634	+ 1478.00	1876.86	3.5e-96	1365	BC021634 Mus musculus, nitrila
gb_hcg:AC084821	+ 1105.00	1357.80	2.9e-67	211771	AC084821 Mus musculus chromo
gb_ro:AF069985	+ 1103.00	1388.07	5.5e-69	4481	AF069985 Mus musculus nitrilas
gb_hcg:AC087229	+ 1083.00	1329.08	1.1e-65	213203	AC087229 Mus musculus chromo
gb_hcg:AC105589	+ 1064.00	1306.45	2.1e-64	178418	AC105589 Rattus norvegicus c
gb_ov:AF284575	+ 968.50	1228.18	4.4e-60	1214	AF284575 Xenopus laevis Nlt pr
gb_hcg:AC017755	+ 715.50	865.19	7.9e-40	131853	AC017755 Drosophila melanoga
gb_in:AC093121	- 715.50	863.04	1.1e-39	173494	AC093121 Drosophila melanoga
gb_in:AC093101	- 715.50	862.77	1.1e-39	174279	AC093101 Drosophila melanoga
gb_hcg:AC006169	+ 715.50	862.73	1.1e-39	190351	AC006169 Drosophila melanoga
gb_in:AE003467	+ 715.50	861.95	1.2e-39	298640	AE003467 Drosophila melanoga
gb_in:AF069989	+ 710.50	857.97	2.0e-39	1521	AF069989 Drosophila melanoga
gb_in:AF069986	+ 702.00	888.24	4.1e-41	1385	AF069986 Caenorhabditis elegans
gb_pl:AF372804	+ 648.50	821.56	2.1e-37	1172	AF372804 Arabidopsis thaliana
gb_hcg:AC010013	+ 632.50	771.43	1.3e-34	34013	AC010013 Drosophila melanogas
gb_ba:AE000934	- 537.50	657.78	2.9e-28	15124	AE000934 Methanodactylum the
gb_pl:SCJUL126W	+ 519.00	653.90	3.7e-28	1307	AE000934 S.cerevisiae chromosome
gb_pl:AF284571	+ 519.00	653.90	4.6e-28	1591	AF284571 Saccharomyces cerevis
gb_ba:AE004861	+ 502.00	615.61	6.3e-26	10466	AE004861 Pseudomonas aerugin
gb_pl:SPC651	+ 502.00	607.79	1.7e-25	25350	AE005572 S.pombe chromosome 1
gb_pr:AF260334	+ 493.00	624.72	2.0e-26	1019	AF260334 Homo sapiens CUX002 m
gb_pl:SCGCD14	- 493.00	615.82	6.1e-26	2792	AE260334 Homo sapiens CUX002 m
gb_ba:SYCSLND	- 490.00	579.47	6.5e-24	110908	AE260334 Homo sapiens CUX002 m
gb_pr:BC020620	+ 487.50	618.19	4.5e-26	965	BC020620 Homo sapiens, Nlt prot
gb_pr:AF284574	+ 485.50	615.67	6.2e-26	965	BC020620 Homo sapiens, Nlt prot
gb_ba:AE005512	- 481.00	600.02	4.7e-25	3180	AF005512 Nostoc commune trna-tl
gb_ro:AF284573	- 481.00	607.34	1.8e-25	1282	AF284573 Mus musculus Nit prot
gb_ba:AB017194	+ 479.50	593.90	1.0e-24	4762	AB017194 Plectonema boryanum C
gb_ro:BC020153	+ 479.00	593.90	1.0e-25	1063	BC020153 Mus musculus, RIKEN C
gb_ba:AP003587	+ 477.50	553.56	2.0e-22	343550	AP003587 Nostoc sp. PCC 7120
gb_ba:AL646071	+ 471.50	550.56	2.6e-22	203050	AL646071 Ralstonia solanaceae
gb_sts:G72919	+ 471.00	598.33	5.8e-25	847	G72919 MARC 2849-2850:1991933517

gb_ba:SM591791 - 444.50 511.59 3.9e-20 340900 ! AL591791 Sinorhizobium me
gb_ba:AJ414158 + 440.50 509.78 4.9e-20 235050 ! AJ414158 Yersinia pestis
gb_ba:AE004129 + 436.50 531.83 2.9e-21 10881 ! AE004129 Vibrio cholerae c
gb_ba:AE005759 - 433.50 526.61 5.7e-21 12748 ! AE005759 Caulobacter cresc
gb_pl:SPAC26A3 + 430.50 513.10 3.2e-20 38149 ! Z69240 S.pombe chromosome

seq_name: gb_pr:AF069987

seq_documentation_block:

LOCUS AF069987 1385 bp mRNA linear PRI 23-JUL-1998

DEFINITION Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.

ACCESSION AF069987

VERSION AF069987.1 GI:3228665

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

17 uargMeValleuallaleeserCysArgThTySerleuserArga 34

|||||
53 GCCAATGGTTTGGCTATATCTTCATGCTG..... 82
34 rgrProArgrLeuGlyPheIleThrArgProPheIleSerLeu 50
83GGCTTCATCCACAGGCTCTCCACAGATTCTGTCTCTT 121
51 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67
122 CTTGTCTGTGAGACTCCGATACCTCAACTCTCAGTACTTGTCTCAGCC 171
67 oArgrProArgrAlaMetAlaIleSerSerSerSerCysGlnLeuProLeu 84
172 CAGGCCAGAGCCATGGCTATCTCTCTCTCTCTCGAATCCTCCCTGG 221
84 AlaAlaValCysGlnValThrSerThrProAspLysGlnIleAsnPhel 100
222 TTGGCTGTGTGCCAGTACATGACGCCAGACAGACAGAACTTTAA 271
101 ThrCysAlaGlnLeuValArgGlnAlaIleArgLeuGlyAlaCysLeuAl 117
272 ACATGTCTGAGCTGGTTCGAGAGCTGCCAGACTGGTCTGCTGGCC 321
117 aPheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGlnThrL 134
322 TTTCCTGCTGAGGCAATTGACTTCATTGACGGACCTGCAGAGAGCG 371
134 euHISLeuSerGlnProLeuGlyGlyLysLeuLeuGlnIleuThrGln 150
372 TACACCTGTCTGACACACCTGGGGGAACTTTGGAGAAATACACCAG 421
151 LeuAlaArgGlnCysGlyLeuThrLeuSerLeuGlyGlyPheHISGlnAr 167
422 CTTGGCCAGGAAATGTGACTCTGTCTGCTGGTGGTTCATGAGCG 471
167 gGlyGlnAspTrpGlnGlnThrGlnLysIleTyrAsnGlyHISValLeuL 184
472 TGGCCAGAGCTGGAGCAGACTGAAATCTAACAATTGTACAGTCTGCG 521
184 euAsnSerLysGlyAlaValAlaIleThrTyrArgLysThrHISLeuCys 200
522 TGAACAGCAAGGGGAGTAGTGGCCACTTACAGGAAGACACATCTGTGT 571
201 AspValGlnIleProGlyGlnGlyPrometCysGlnSerAsnSerThrWe 217
572 GACGTGAGATTCACAGGCGAGGGCTATGTGAAAGCAACACTGTACAT 621
217 tProGlyProSerLeuGlnSerProValSerThrProAlaGlyLysIleG 224
622 GCCTGGGCCAGCTTGAGTCACTGTCAACACACAGCAGGCAAGATTG 671
234 IyLeuAlaValCysTyrAspMetArgPheProGlnLeuSerLeuAlaLeu 250
672 GTCTACTCTGTGCTATGACATCGGCTCCCTAACTCTCTCGGATTC 721
251 AlaGlnAlaGlyAlaGlnIleLeuThrTyrProSerAlaPheGlySerI 267
722 GCTCAAGCTGGAGCAGAGATTACTTACCTTCCCTTCAGCTTTTGATCCAT 771
267 eThrGlyProAlaHISTrpGlnValLeuLeuArgAlaIleGlnT 284
772 TACAGGCCAGGCCACTGGGAGGTGTGTGGGGGCCCTGTCTATCGAA 821
284 hrcLncYstyValValAlaAlaIleGlnCysGlyArgHISHisGlnLys 300
822 CCCAGTGTATGTAGTGCAGCAGCAGCATGTGGAGCCACCACTGGAAAG 871
301 ArgAlaSerTyrGlyHISerMetValAlaAspProTrpGlyThrValVa 317
872 ACAGCAAGTTATGGCCAGCATGTGTGAGACCCCTGGGGAACAGTGGT 921
317 lAlaArgCysSerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeu 334
|||||

922 GGGCCGCTGCTGAGAGGGCCAGGCTCTGCTTGGCCCGAATAGACTCA 971
334 sntYrLeuArgrGlnLeuArgrArgHISLeuProValPheGlnHISArgArg 350
972 ACTATCTGCAGAGTGTGGCCGACACTGCTGCTGTTCAGACGCCAGG 1021
351 ProAspLeuTyrGlyAsnLeuGlyHISProLeuSer 362
1022 CCTGACCTCTATGCAATCTGGGTCAACCACTGTCT 1057
seq_name: gb_hcg:AL354714_4
seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (5 of 7) of AL354714 from base 400001 (AL354714 Homo sapiens chromosome
alignment_scores:
Quality: 1538.50 Length: 888
Ratio: 4.262 Gaps: 9
Percent Similarity: 40.653 Percent Identity: 40.315
alignment_block:
US-09-357-675c-25 x AL354714_4 ..
Align seg 1/1 to: AL354714_4 from: 1 to: 110000
1 ProLeuAlaAlaAla**LeuAlaProAspArgProProAspArgThrIle 17
|||||
103174 CCACGCTGCGCGGCTGTGCTGCCAGACCGCCCTCCGATCGGACCT 103223
17 uArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 34
|||||
103224 GCGAATGTTTGGCTATATCTTATGATGAGACTACTCCATATCCGATC 103273
34 rgrPro..... 35
103274 GGCCGGGGTGAATCCACCTCGCGTCTTTAACTGTGTAACAGAGATGC 103323
35 35
103324 TGCCTCTGGAGAGCGCGGGAGGAGCGGCCAACATGAGCGGGCGCGG 103373
35 35
103374 AGGCTGAGAGCGCGGGCGGCTTGGGGCTTGAGTCTTTGCCCTTGC 103423
35 35
103424 CCACAGGAGGGGTGGAGACGAGAGAGGTTGAATTTCCTCCCTCGGAGA 103473
35 35
103474 TTCTGTGAAGGGAATATGCTTCAGTCAAGTAAGCTGGCAAGTGC 103523
35 35
103524 ACAGTCAAGAGAGAGTCTTGGAAACCAAGAGATGTTCCCGGAGATGA 103573
35 35
103574 CTTTGGACTGGGGAAGCTTTGTCAAGAGAAAGAGCTTCAAGTTAATG 103623
35 35

103624 TGAGATCATTTGGAAGTTGAACATATTCAGGGGGGAGAGAGAGCCCTTTCT 103673
 35 35
 103674 CCATCTCCCACTGTATGTCTGTCTCCCTTAGCCCTAATTTCTATATAA 103723
 35 35
 103724 GAGCGTTTGAGACTAGTTTCTTACTGGCAGGGCCCTGATTGAGCTCA 103773
 35 35
 103774 CAGTATAAGAGAGAGAGTCAAGAGATCTAGAAATTTATTTGCTTGAAGA 103823
 36 ArgLeuGlyPheIleThra 42
 103824 GGGGTGTGTCTCCATATCTCACCCTTCCCTGAGGCTGGCTTCATACCA 103873
 42 rGProH3sarGpHeuSerLeuLeuCysProGlyLeuArgIlePro 58
 103874 GGGCTCTCAGACATTCCTGTCTCTGTCTGTCTGTGACTCGGATACCT 103923
 59 GlnLeuSerValLeuCysAlaGlnPro 67
 103924 CAACCTCTCAGTACTTTGTGCTCAGCCAGSTAACAGSTTTTGTGTGTCC 103973
 67 67
 103974 TCAGTCCCTGGCAGCTTAGATGCTCAGTTGTGTAATGATAGTGGAGAC 104023
 67 67
 104024 ACAGAGGTCAACTATTCACACATTTGATTTGTGAGCCCTACTAGCCCT 104073
 67 67
 104074 GGGTCACAGTGGCCCTGTAGAGCATGATCAAAAGAAAGTCCACTTTCT 104123
 67 67
 104124 GCCTTCACCTTGACCCCTTAGCATTAATTTGCTCTCCCTGCTATGA 104173
 68 ArgProAlaIleMetAlaIle 74
 104174 AATCTGAGATCTCGCTATGCTGTTCACAGGCCAGAGCCAGGCTATC 104223
 75 SerSerSerCysGlnLeuProLeuValAlaValCysGlnValThrse 91
 104224 TCCTTCTCCCTCGGAGACTGCCCTGGTGTGTGTGTGCCAGGTAACATC 104273
 91 rThrProAspLysGlnGlnAsnPheLysThrCysAlaGlnLeuValArg 108
 104274 GAGCGAGACCAACACAGAACTTTAAACATGTGCTGAGCTGGTTGAG 104323
 108 LysAlaIleArgLeuGlyAlaCysLeuAlaPheLeuProGlnAlaIlePheAsp 124
 104324 AGGCTCCAGACTGGGTGCTGCTGCTTCTGCTGAGGCAATTTGAC 104373
 125 PheIleAlaArgAspProAlaGlnThrLeuHisLeuSerGlnProLeuG 141
 104374 TTGATTGCAAGGAGCCCTGAGAGACGCTACACCTGTCTGACACACTGG 104423
 141 yGlyLysLeuLeuGlnLysThrGlnLeuAla 152
 104424 TGGGAAACTTTTGAAGAATACACCCAGCTTGCCAGGTATCAGGAATA 104473
 152 152
 104474 GCGAGGAGAGGTAGATCTTTGTTGGACAGTGTCCCTGGGTGCCAGAT 104523
 152 152

104524 ATGAGGTAGAGCCTTGAGAAAGTCAAGTGAAGAGTGTGTCAGTTCCTTCC 104573
 153 ArgGluCysGlyLeuThrLeuSerLeuGlyPheHisGlnArg 168
 104574 CCCAGGGAATGTGAGACTGTGGCTGTCCCTGGGTGCTTCCATGAGCGTG 104623
 168 LysGlnAspTrpGlnGlnThrGlnLysIleTyrAsnGlyHisValLeuLeu 184
 104624 GCCAGACTGGGAGCAGACTCAGAAAATCTACAAATTTGTCTCAGTGTCTG 104673
 185 AsnSer..Lys 187
 104674 AACAGCAAGGTGAGACTTATTAACCTTAGCCTGCTCTCCATGCT 104723
 187 187
 104724 CTTCACCTAGAAATTCACGAATTTGTTCTCAACTTATTTCTTGACC 104773
 187 187
 104774 CAAGATTTAGGGGTGCTCTACTTCACTTCTAGCTTAACTAATATCTC 104823
 187 187
 104824 CTCCTGGAGAGAGTAGCAAGGCTCTAGAACACACAGACTGATTTCC 104873
 188 GlyAlaValAlaIleThrTyrArgLysThrHisLe 199
 104874 TTCTTTCTTACTGTAGGGGAGTAGTGCCACTTACAGGAAGACACTCT 104923
 199 uCysAspValGluIleProGlyGlnGlyProMetCysGlnSerAsnSer 216
 104924 GTGTGACGTAGAGATTCACAGGACAGGGCCCTATGTGTAAGACACTCTA 104973
 216 hMetProGlyProSerLeuGlnSerProValSerThrProAlaGlyLys 232
 104974 CCATGCTGGGCCAGTCTTGTGATCACCCTGTACAGCACACAGCAGGCAAG 105023
 233 IleGly..Leu 235
 105024 GTAGAGATGTTGAAGAGATGAGGAGGGAACAGAAATACTTTGAACTGG 105073
 235 235
 105074 CAGTAGAGATAGAAAGCCCTAAGAGAGGGGTAAATGAAATATAGTAG 105123
 235 235
 105124 ATGCTGTGACAAACAGACAGCAAGAAAGTACTAAGTAGGCTTTTCATT 105173
 236 AlaValCysTyrAspMetArgPheProGlnLeuSerL 248
 105174 CCAAGATTGCTGTAGCTGTGTGCTATGACATGCGGTTCCCTGAACCTCTC 105223
 248 euAlaLeuAlaGlnAlaGlyAlaGlyLysLeuThrTyrProSerAlaPhe 264
 105224 TGGATTGGCTCAAGCTGAGCAGAGAACTTACTTACCTTCAGCTTTT 105273
 265 GlySerIleThrGlyProAlaHisTrpGlnVal..Leu 276
 105274 GGAATCAATTACAGGCCAGCCCACTGGAGAGTAAAGATGATGCCCTTTTAA 105323
 276 276
 105324 AACATAAGGCGTTTCTTAAACCTATCTTCCCTTGCCCTTACACAGT 105373
 276 276
 105374 TAAATTCCTTCCCTTTCACACTAATGGAAGAACTCATTTCCCCAGATATT 105423
 276 276
 105424 TCTCTCATGAATAGTTAAATAGTAATCATTTCTTAGCAATTACAAA 105473

276 276
105474 TAGTCACATGGGTAGATTGCTCTGTAATGTCCCTCACCTGTCACTTCCC 105523
276 276
105524 CACTATTTCCTACATGTACTTAAGTAGACACACATCTCATGCCAGGTGT 105573
277 .. LeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAla 292
105574 TGTGTGGGGGCCGCTGCTATCGAAACCAAGTGTATGTAGTGGACAGCA 105623
293 GlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetVal 309
105624 CAGTGTGACGCCACCATGAGAGAGCAAGTTATGGCCACAGCAATGGT 105673
309 ValAspProTyrPglYthrValValAlaArgCysSerGluGlyProGlyL 326
105674 GGTAAGCCCTGGGGACAGTGGTGGCCGCTCTGAGGGGCCAGGCC 105723
326 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 342
105724 TCTGCTTGCCTCCGAATAGACCTCACTATCTGCGACAGTTGCCCGACAC 105773
343 LeuProValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyH 359
105774 CTGGCTGTCTCCAGACCGCAGCGCTGACCTCTATGGCAATCTGGGTCA 105823
359 sProLeuSer 362
105824 CCCACTGTCT 105833
seq_name: gb_hfeg:AL354714_5

seq_documentation_block:

WPCOMMENT

Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_1 1 110000
AL354714_2 110001 210000
AL354714_3 210001 310000
AL354714_4 310001 410000
AL354714_5 410001 510000
AL354714_6 510001 610000
AL354714_7 610001 699820
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome 1 c

alignment_scores:

Quality: 1538.50 Length: 888
Ratio: 4.262 Gaps: 9
Percent Similarity: 40.653 Percent Identity: 40.315

alignment_block:

US-09-357-675c-25 x AL354714_5 ..

Align seg 1/1 to: AL354714_5 from: 1 to: 110000

1 ProLeuAlaAlaAla***LeuAlaProAspArgProAspArgThrLe 17
|||||
3174 CCACCTGCGCGCGGCTTCTGCTCCAGACCGCCCTCCGATCGGACCT 3223
17 uArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 34
|||||
3224 GCGAATGGTTTGGCTATATCTTCATGTAGACCTACTCCCTATCCCGTC 3273
34 rgrPro..... 35
|||||
3274 GCGCGGGTGAATCCACCTGCGGTGCTTTAACTGTGTACAGAGATGC 3323
35 35

3324 TGCCTCTGGAGAGCGGGGAGGAGCGGCCAACCTGAGCGGGCGGG 3373
35 35
3374 AGGGTGGAGGGCGGGCGGGCTTGGGGCTGGGTTCTTTGGCCCTTGC 3423
35 35
3424 CCACAGAGGAGGGTGGAGAGCAGAGAGGAGTGAACCTTCCCTGCGAGA 3473
35 35
3474 TTCTGTGAAGGGGAATATGCTTGCAGTCACTAAAGCTGCGCAATGC 3523
35 35
3524 ACAGTCAAGAGAGAGTCTTGGGAAAACAGAGATAGTTCCCGAGATGA 3573
35 35
3574 CTTTGGACTGCGGAAACGTTTGTACAGAGAAAGAGCTTCACTTAAAG 3623
35 35
3624 TGAGATCAATTGGAAGTTGAATTCAGGCGGCGAGACAGCCCTTCT 3673
35 35
3674 CCATCTCCACCTGATGGTCTTGTCCCTTAGCCTAATTTCTTAA 3723
35 35
3724 GAGCGTTTGAAGACTAGGTTTCTTACTGGCAGGCGCCGTATCAGGCTCA 3773
35 35
3774 CAGTATTAAGAGAGAAAGTCAAGATCTAGAAATTTATGCTTGAGAA 3823
36 36
3824 GGGGTGGTGTCTCTCATATCTCACTTCTCAGGCTGGGCTTATCAGCA 3873
42 rgrProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIlePro 58
|||||
3874 GGCTCTCAGAGATTCCTGCTCTGTCGTGTCGATCCGATACCT 3923
59 GlnLeuSerValLeuCysAlaGlnPro..... 67
|||||
3924 CACTCTCAGTACTTGTGTCTCAGCCAGGTACACGTTTGTGTGTCC 3973
67 67
3974 TCAGTGCCTGGCACTTAGATGCTCACTTGTTAATGATGATGGGAGAC 4023
67 67
4024 ACAGAGTGTCAACTATCCACACATTTGATTTGGTGAAGCCCTACTAGCCCT 4073
67 67
4074 GGGTCAAGCTGCCCTGTAGAGACATGATCAAAAAGAAATGCCACTTCTCT 4123
67 67
4124 GCCTTCCACTTGCAACCCCTTAGCATTAATTTGCTTCCCTGTGCTATGA 4173
68 68
4174 AATCTGAGATCTCGCTATGCTGTTCACAGGCCCGCAGACCATGCTATTC 4223
75 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSe 91
|||||
4224 TCTCTTCTCTCGGAACTGCCCTGTGGTGTGTGTCCAGGTATACATC 4273

```

91 rThrProAspLysGlnGlnAsnPhelYsThrCysAlaGluLeuValArg 108
   |||||
4274 GAGCCGACGACAGACAGACAGAACTTTAAACATGTGCTGAGCTGGTTCCAG 4323
108 lualalaArgLeuGlnAlaCysLeuAlaPheLeuProGluAlaPheAsp 124
   |||||
4324 AGCGTGCAGAGCTGGTGCCCTGGCTGCTTCTGCTGAGGCAATTTGAC 4373
125 PheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGlu 141
   |||||
4374 TTCATTCACGCGACCTGCAGAGACGCTACACCTGCTCGAACCACTGGG 4423
141 YGlyLeuLeuGluGluGluThrGlnLeuAla..... 152
   |||||
4424 TGGGAACCTTTGGAGAGATACCCAGCTTGCCAGGTATCAGGGAATA 4473
152 ..... 152
4474 GCGAGGAGAGGTAGAACTTTGTTGGACAGTGTCCCTGGGTGCCAGAT 4523
152 ..... 152
4524 ATGAGGGTAGACCTTGAGAACTCAGTGAAGAGTGTGTCAGTCCCTTCC 4573
153 .... ArgGluCysGlyLeuThrPheLeuSerLeuGlyGlyPheHisGluArg 168
   |||||
4574 CCCCGAGGAATGTGACTGTGGCTGTGCTGGGTTCCTGATGAGCTGTG 4623
168 lYglnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeu 184
   |||||
4624 GCCAGACTGGGAGCAGACTCAGAAATCTACAAATGTGCAGCTGCTGTG 4673
185 AsnSer..Lys..... 187
   |||||
4674 AACGCAAGAGGTGAGACTTTATAACCTTTAGCCTGTCCATCTCCATGCT 4723
187 ..... 187
4724 CTTCACCTAGATTCCTCAGAAATGTTTCTCACTTTATTTCTTGACC 4773
187 ..... 187
4774 CAAGATTTAGGGGTGCTCTACTTCACTTCTAGCCCTATAACTATCTC 4823
187 ..... 187
4824 CTCCTTGGAGAGTAAAGCAAGGCTTCTAGAACAACAGCACTGATATTC 4873
188 .....GlyAlaValAlaIleThrTyrArgLysThrHisIle 199
   |||||
4874 TTCCTTTCTTACTGTAGGGGAGTGTGGCCACTTACAGGAAGACACATCT 4923
199 uCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 216
   |||||
4924 GTGTAGAGTATAGATTCAGGGGAGGGGCTATGTGTAAGCAACTCTTA 4973
216 hMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLys 232
   |||||
4974 CCATGCTGGGCGCCAGTCTTGATCACCCTGCAGACACAGCAGGCAAG 5023
233 lIleGly..Leu..... 235
   |||||
5024 GTAGAGATTGTGAAGATGAGGAGGAGGGAACAGAAATACTTTGAATGG 5073
235 ..... 235
5074 CAGTAGAGATAGAAAGCCCTAAGAGAGGGGTAATGAAATATGACTAG 5123
235 ..... 235
5124 ATGCTGTGACAAACAGACAGGAAGACTACTAAGTAGGCTGTTTTCATT 5173

```

```

236 .....AlaValCysTyrAspMetArgPheProGluLeuSerL 248
   |||||
5174 CCAGATTGGTGTAGCTGTCTGTCTATGACATGCGGTTCCTCCCTGCTCTC 5223
248 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 264
   |||||
5224 TGGCATTTGCTCAAGCTGGAGCAGAGATATCTACTATCTTCACCTTTT 5273
265 GlySerIleThrGlyProAlaHisTrpGluVal..Leu..... 276
   |||||
5274 GGATTCATTTACAGGCCAGCCACCTGGAGGATGAGATGATGCTTTTAA 5323
276 ..... 276
5324 AACATAGGGCCTTTTCTTAACCTCATCTTCCCTTGCCCTACAGT 5373
276 ..... 276
5374 TAAATTCCTCCCTTCCACCTAATGGGAAAACTCATTCGCCAGATAT 5423
276 ..... 276
5424 TCTCTCATGAATAGTTAAATAGTAAATCATTCCTAGGCAATTACCAAA 5473
276 ..... 276
5474 TAGTCACATAGGTAGATTGCTGTATATGTCCTCACCCTGCACTTCC 5523
276 ..... 276
5524 CACTATTTGCTACATGTACTTAAGTGAACACACATCTCATGCCAGGTGT 5573
277 ..LeuArgAlaArgAlaIleGluThrGlnCysTyrValAlaAlaAla 292
   |||||
5574 TGCTGGGGGCCGTGCTATCGAACCAGTGTATGTGTGTGGCCACACACA 5623
293 GlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetVa 309
   |||||
5624 CAGTGTGAGCCGACCATGACAGAGAGACAGTTATGGCCACAGCATGTGT 5673
309 lValAspProTrpGlyThrValAlaAlaArgCysSerGluGlyProGlyL 326
   |||||
5674 GSTAGACCCCTGGGGAACAGTGTGGCCCGCTGCTCAGGGGCGCAGGCC 5723
326 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgHis 342
   |||||
5724 TCTGCTTGGCCCGAATGACCTCACTATCTGCGCAGCTTGGCGGACAC 5773
343 LeuProValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHI 359
   |||||
5774 CTGCTGTGTTCAGACACCGCAGGCTTACCTCTATGGCAATCTGGTCA 5823
359 sProLeuSer 362
   |||||
5824 CCCACTGTCT 5833
seq_name: gb_hhg.AL590651
seq_documentation_block: 167863 bp DNA linear HTG 18-AUG-2001
LOCUS AL590651
DEFINITION Homo sapiens chromosome 1 clone RP11-137A12, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL590651
VERSION AL590651.4 GI:13990236
KEYWORDS HTG: HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 167863)
AUTHORS Harrison,E.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,

```



```

35 ..... 35
151476 CAGTATTAAGAGAGAACTCAGAGATCCTAGAATTATTGCTTGAGAA 151427
36 ..... ArgLeuGlyPheIleThrA 42
151426 GGGGTTGGTTCCTCATTCTCAGCTTCCTCAGAGCTGGGCTTCATCACC 151377
42 rPrProHISarGpHeLeuSerLeuLeuCysProGlyLeuAlaGlyIlePro 58
151376 GGGCTCCTCAGACAAATTCGTGCTTCGTCTGTGCTGACATCCGATACCT 151327
59 GluLeuSerValLeuCysAlaGlnPro ..... 67
151326 CAACCTCTCAGTACTTGTGCTCAGCCAGGTAACAGCTTTTGTGTGTCC 151277
67 ..... 67
151276 TCAGTGCCTGGACCTTAGATGCTCAGTTGGTAAATGAGATGGGAGAC 151227
67 ..... 67
151226 ACAGAGTGTCAACTATCCACACATTTGATGGTGAGCCCTACTAGCCCT 151177
67 ..... 67
151176 GGGTCACGTCGCCCTGTAAAGCATATCAAAAGAAATCCAGCTTTCCT 151127
67 ..... 67
151126 GCCTTCACCTTGCACCCCTTAGCATTTAAATTGCTCCCTGTGCTATGA 151077
68 ..... ArgProAlaMetAlaIle 74
151076 AATCTGAGATCCTGCCTATGCTGTTCACAGGCCAGAGCCATGCGTATC 151027
75 SerSerSerCysGluLeuProLeuValAlaValCysGlnValIlePhe 91
151026 TCCTCTCTCCTCGCGACACTGCCCTGGTGCTGTGTGCCAGGTAACATC 150977
91 rThrProAspLysGlnGlnAsnPhelSerThrCysAlaGluLeuValArg 108
150976 GAGCCAGACAAAGCAACAGAACTTTAAACATGCTGCTGAGCTGTTCCAG 150927
108 LuAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAsp 124
150926 AGGCTGCCAGACTGGGCTGCTGCTGCTTTCCTGCCAGGCAATTTGAC 150877
125 PheIleAlaArgAspProAlaGluThrIleuHisLeuSerGluProLeuG 141
150876 TTCATTGCAAGGGACCCTGCAGAGAGCTACACCTGCTTAACACACTGGG 150827
141 yGlyLysLeuLeuGluGluIuThrGlnLeuAla ..... 152
150826 TGGGAACCTTTTGAAGAATACACCCAGCTTGCAGAGTATCAGGAATA 150777
152 ..... 152
150776 GCGAGGAGAGGTAGATCTTTGTGGACAGTCCCTGGGTTGCCAGAT 150727
152 ..... 152
150726 ATGAGGGTGAAGCCTTGAGAACTGCAAGAGTTGTCAGTGCCTCTCC 150677
153 .... ArgGluCysGlyLeuThrPheLeuSerLeuGlyGlyPheHisGluArg 168
150676 CCCAGAGGAATGTGCACTGTGGCTGTCTGCTGGGTTCATGATGAGGCTG 150627
168 LysGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeu 184
150626 GCCAAGACTGTGAGCAGAGACTCAGAAAATCTACAAATTGTACAGTGTCTG 150577

```

```

185 AsnSer..Lys ..... 187
150576 AACAGCAAAAGGTAGACTTTATACCTTTAGCCTGTGCTTCCATGCT 150527
187 ..... 187
150526 CTTCACCTAGATTCTCCAGAAATTGTTTCAACTCTATTATTCCTTGAC 150477
187 ..... 187
150476 CAAGATTAAAGGGGTGCTCTACTTCACTTCTAGCCCTATTAACATATC 150427
187 ..... 187
150426 CTCCTTGGAGAGTAAGCAAGCTTCTAGAACACCAACGACATATATCC 150377
188 ..... GlyAlaValAlaIleThrTyrArgLysThrHisLe 199
150376 TTCTTCTTACTGTAGGGCAGTAGTGCCACTTACAGAAACACATCT 150327
199 uCysAspValGluIleProGlyGlnGlyPrometCysGluSerAsnSer 216
150326 GTGTGACGTAGAGATTCCAGGGCAGGCTTATGTGAAAGCAACTCTA 150277
216 hMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLys 232
150276 CCATGCTTGSGGCCAGCTTGTGAGTCACTGTACACACCAACGACGCAAG 150227
233 IleGly..Leu ..... 235
150226 GTAGAGATTGTAAAGAGATGAGGAGGGAACAGAAATCTTGAATCG 150177
235 ..... 235
150176 CAGTAGAGATAGAAAGCCCTAAGAGAGGGGTAAAGAAATATGACTAG 150127
235 ..... 235
150126 ATGCTGTGACAAACAGACGAGGAAGACTACTAAGTAGGCTGTTTTCAT 150077
236 ..... AlaValCysTyrAspMetLargPheProGluLeuSerL 248
150076 CCAGATTGCTGTAGCTGTGTATGACATGCGGTTCCGAACTCTCTC 150027
248 euAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPhe 264
150026 TGGCATTGGCTCAAGCTGGAGAGAGATTACTTACCTTCTTCCAGCTTT 149977
265 GlySerIleThrGlyProAlaHisTrpGluVal..Leu ..... 276
149976 GGATCCATTACAGGCCCAAGCCACCTGGAGGTAAGATGATGCCTTTAA 149927
276 ..... 276
149926 AACATAGGGCCTTTTCTTAACCTCATCTTCCCCCTTGGGCCCTACAGT 149877
276 ..... 276
149876 TAAATTCCTTCCCTTCCACCTAATGGAAGAACTCATTCGCCAGATAT 149827
276 ..... 276
149826 TCTCTCATGAAATGTTAAATAGTAATCATTTCTTAGGCAATTACAAAA 149777
276 ..... 276
149776 TAGACCAATGGGTAGATGTGTGTATGATGCCCTGACCTGTGCTCC 149727
276 ..... 276
149726 CACTATTGTCTACATGACTTAAGTGAACACATCTCATGCCCAGGTGT 149677
277 ..LeuArgAlaArgAlaIleGluThrGlnCysTyrValAlaAlaAla 292

```

```

149676 TTCTGGGGCCCGCTATCGAAACCACTGCTATGAGCAGCA 149627
293 GINCYGAIARIGHISIGLULYSARGALASERTYRGLVHISSEMETYA 309
149626 CAGTGTGGACGCGACCATGAGAGCAAGTTATGGCCACGCAWTGGT 149577
309 IVALASPPRTPRGlyThrValValAlaIARGCYSSErGLUGlyProGLY 326
149576 GTTAGACCCCTGGGAGACAGTGTGGCCGCTGCTGAGGGCCAGGCC 149527
326 euCysLeuAlaArgIleAspLeuAsnTYrLeuArgGlnLeuArgArgHis 342
149526 TCTGGCTTGCAGACCTCAACTATCTGAGACAGTTGCCCGCACAC 149477
343 LeuProValPheGlnHisArgArgProAspLeuTYrGlyAsnLeuGlyH 359
149476 CAGCTGTGTTCCAGACCGCAGGCTGACCTATAGCAATCTGGGTCA 149427
359 sProLeuSer 362
149426 CCCACTGTCT 149417
seq_name: gb_pr:AL591806
seq_documentation_block:
LOCUS AL591806 200822 bp DNA linear PRI 30-JAN-2002
DEFINITION Human DNA sequence from clone RP11-544M22 on chromosome 1, complete
sequence.
ACCESSION AL591806
VERSION AL591806.16 GI:18476709
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 200822)
AUTHORS Harrison, E.
TITLE Direct Submission
JOURNAL Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail requests: clonesquest@sanger.ac.uk
humquerry@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:117902927.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP11-544M22 is from the library RPI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/Dacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-544M22. The true
left end of clone RP11-137A12 is at 155538 in this sequence. The
true right end of clone RP11-381D2 is at 145015 in this sequence.
Location/Qualifiers

```

```

source 1. .200822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-544M22"
/clone_1lb="RPI-11.2"
10395. .10397
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
44098. .44107
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
59585. .59604
/note="Sequence from overlapping clone RP11-381D2
(AL162592). Assembly confirmed by restriction digest."
90797. .90883
/note="Single clone region. Reads generated from a
transposon library derived from a single pUC clone.
Restriction digest data confirm the assembly."
90861. .90883
/misc_feature /note="Sequence from uni-directional dGTP big dye
terminator reads only."
BASE COUNT 55215 a 49052 c 46183 g 50372 t
ORIGIN
alignment_scores:
Quality: 1538.50 length: 888
Ratio: 4.262 gaps: 9
Percent Similarity: 40.653 Percent Identity: 40.315
alignment_block:
US-09-357-675c-25 x AL591806 ..
Align seg 1/1 to: AL591806 from: 1 to: 200822
1 ProLeuAlaAlaAla**LeuAlaProAspArgProAspArgThrLe 17
|||||
172105 CCACCTGCTGCGGCGCTCTGCTCCAGACCGGCTCCGAGTCGACCT 172154
17 UArgMetValIleuAlaIleSerSerCysArgThrTYrSerLeuSerArg 34
|||||
172155 GCGAAAGGTTTGGCTATCTTCATGAGAGCACTCCATCCGCTC 172204
34 rgPro.....
|||||
172205 GCGCGCGTGCAATCCCACTCGGGTCTTAACTTGTGAACAGAGATGC 172254
35 .....
172255 TGCTGTGGAGAGGCGGGAGGAGGACGCAACTGAGCGGGCGGG 172304
35 .....
172305 AGGCTGAGAGGCGGGCGGGCTGGGGCTTGCTTCCCTTCCCTTGC 172354
35 .....
172355 CCACCAAGGAGGGGTGGAGACGAGAGAGGTGAACCTTCCCTCGGAGA 172404
35 .....
172405 TTCTGTGAAGGAGGAATATGCTTCGAGTCAGTAAGCTGCGCAAGTGC 172454
35 .....
172455 ACAGTCAGAGAGAGAGCTTGGGAAACCAAGATATGTCGCGAGATGA 172504
35 .....
172505 CTTTGTGACTGCGGAACGTTGTGAGAGAGAAAGAGCTTCACTTAATG 172554
35 .....
Location/Qualifiers

```



```

276 ..... 276
174405 TACTCACAAATGGGTAGATTGGTCTGTAATGTCCTCCTCAGCTGACTTCCC 174454
276 ..... 276
174455 CACTATTGCTACATGACTTAAGTGAAACACATCTCATGCCAGGTGT 174504
277 .. LeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAla 292
174505 TCGTCGGGGCCCGCTGATCGAAACCCAGTGTATGAGGACAGCA 174554
293 GlnCysGlyArgHisHisGlnLysArgAlaSerTyrGlnHisSerMetVal 309
174555 CACTGTGGAGCCACCATGAGAGAGCAAGTTATGCCACAGCATGTT 174604
309 ValAspProTyrGlyThrValValAlaArgCysSerGlnGlyProGlyL 326
174605 GGTAGACCCCTGGGGAACTGGTGGCCCTGCTGTGAGGGGCCAGGCC 174654
326 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 342
174655 TCTGCTTGGCCGAATAGACCTCACTATCTGGACAGTTGCCGCCGAC 174704
343 LeuProValPheGlnHisArgArgProAspLeuTyrGlnLysLeuGlyH 359
174705 CTGCGCTGTTCAGACGCGCAGCGCTGACCTGATGCAATCGGTCA 174754
359 sProLeuSer 362
174755 CCCACTGCTT 174764
seq_name: gb_pr:AF069984
seq_documentation_block:
LOCUS AF069984 4079 bp DNA linear PRI 23-JUL-1998
DEFINITION Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced
product, complete cds.
ACCESSION AF069984
VERSION AF069984.1 GI:3242977
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4079)
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrilase and flht homologs are encoded as fusion proteins in
drosophila melanogaster and caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
JOURNAL
MEDLINE 98337986
TITLE 2 (bases 1 to 4079)
Pekarsky,Y., Campigilio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
source
1..4079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q21-q22"
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3569)
/gene="NIT1"
/note="alternatively spliced"
/product="nitrilase homolog 1"
635..3569
gene

```

```

exon /gene="NIT1"
635..680
/note="alternatively spliced"
CDS /number=1
join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,
2609..2734,3001..3569)
/gene="NIT1"
/note="NIT1; alternatively spliced"
/codon_start=1
/product="nitrilase homolog 1"
/protein_id="AAC39901.1"
/db_xref="GI:3242978"
/translacion="MLCFITRPPHRLSLCPGLRIPOLSVLCAQPRPRMAISSSC
ELPLVAVQVSTPDKQONFKTCAELVEARLGLACLFLEAFDFIARDAETLHS
KPLDGKLEEVYTLARREGIMLSGGFHERODMPOKIKYCHVLSKGAAYATYR
KTHLCDVEIRPGCPMPCESNSTMPGSLSPSTPRAGKIGLAVCYDMRPPESLALAA
GAELITVPSAFSGSITGPAHWEVLLKRALETQCYVAAAGRHHEKRASTGHSNVD
FWGITVVARCSSEPGICLARIDLVTLROLRHLPVQHRHPPDLYGNLGHPLS"
681..711
/gene="NIT1"
/note="1A; alternatively spliced"
712..741
/gene="NIT1"
/note="1B; alternatively spliced"
940..1053
/gene="NIT1"
/note="1C; alternatively spliced"
1288..1383
/gene="NIT1"
/number=2
/number=2
1636..1890
/gene="NIT1"
/number=3
2011..2114
/gene="NIT1"
/number=4
2321..2454
/gene="NIT1"
/number=5
2609..2734
/gene="NIT1"
/number=6
3001..3569
/gene="NIT1"
/number=7
BASE COUNT 940 a 1082 c 1043 g 1012 t 2 others
ORIGIN
alignment_scores:
Quality: 1525.50 Length: 889
Ratio: 4.226 Gaps: 10
Percent Similarity: 40.607 Percent Identity: 40.270
alignment_block:
US-09-357-675C-25 x AF069984
Align seg 1/1 to: AF069984 from: 1 to: 4079
1 ProLeuAlaAlaAla**LeuAlaProAspArgProProAspArgThrIle 17
||||| 17
604 CCACGCTGCTGGCGGCTTGTGCTCCAGACGCGCCGCGGATCGGACCT 653
17 uArg.MetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 33
||||| 33
654 GCCAATGTTTGGCTATATCTTCAATGAGACCTACTCCATCCCGT 703
34 ArgPro..... 35
||||| 35
704 CGGCGCGGTGAATCCACCTCGGCTTTAACTGTGTAAACAGAGATG 753
35 ..... 35

```

754	CTGCCCTTGGGAGAGCGGGAGGGACGGGCCCACTGGAGCGGGCGCGG	803
35	35
804	GAGGGTGGAGGCGGGGGCGGGCTTGGGGCTTGGGCTCTTTGGCCCTTG	853
35	35
854	CCACACAGGAGAGGGTGGGAGACGAGAGAGGGTGAACCTTCCCTGGGAG	903
35	35
904	ATTCTGTGAAGGGGAAATATGCTTCAGAGTCAGTAAAGCTGCCAAGTG	953
35	35
954	CACAGTCAGAGAGAGAGTCTTGGGAAACCAAGATAGTCCGGAGATG	1003
35	35
1004	ACTTTTGGACTGGGAAACGTTTGTACAGAGAAAGAGGCTTCAGTTTAAT	1053
35	35
1054	GTGAGATCATTGGAACTTGAACATTACAGGGGGGAGAGAGGCCCTTTC	1103
35	35
1104	TCCATCTTCCCACTGATATGTTCTGCTCCCTTACGCTATATTTCTTATTA	1153
35	35
1154	AGAGCGTTTGAAGACTAGTTTTTACTGTGAGAGGCCCTGATTCAGGCTC	1203
35	35
1204	ACAGTATAAGAGAGAGAGTCAAGAGATGCTGAAATTTATTGCTTGAAGA	1253
36	41
1254	AGGGGTGGTGTCTCATATCTACACCTTCCTCAGGCTGGGCTTCATCAC	1303
42	ATGPRORHISATGPhenylserineLeuysProGlyLeuAlaGlyLePr	58
1304	AGGCGTCTCAGAGATCTGTCCTTCTGTGTCTGTGACATCCGGATACC	1353
58	ocLeuSerValLeuCysAlaGlyPro.....	67
1354	TCACACTGTGACTTGTGTGTCACAGCCACAGTAACAGCTTTGTGTGTCT	1403
67	67
1404	CTCAGTGCCTGGCACTTAGATGTCATGTTTGTTAATGATAGTGGGAGA	1453
67	67
1454	CACAGAGTGTCAACTATGCACACATTTGATTGGTGAAGCCTTACTAGCCC	1503
67	67
1504	TGGGTCAACGTGCCCTGTAAAGACATGATCAAAAGGAATCCAGATTTC	1553
67	67
1554	TGCCTTCCACTTGGACCCCTTAGCATTAATTTGCTTCCCTGTGCTATG	1603
68	74
1604	MAATCGAATATCTGSCATATCTGTTCACAGGCCCAAGCCATGCGTAT	1653
74	eSerSerSerCysGlyLeuProLeuValAlaValCysGlyValThrS	91

1654	CTCCTCTTCTCTCTCGAACCTCCCTCGTGCTGTGGCCAGGTAACT	1703
91	erThrProkspLysGlnGlnAsnPhelysrhCysAlaGluLeuValArg	107
1704	CGAGCGCAGCAAGCAACAGAACTTTAAACATGTGTGAGCTGTTTCA	1753
108	GluAlaAlaArgLeuGluYlaCysLeuAlaPheLeuProGluAlaIleAs	124
1754	GAGGCTCCAGACTGGGTGCTGCTGGCTTCCGCTGGAGGATTTGA	1803
124	pPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeu	141
1804	CTTCATTCACGGAGGACCTCTGCAGAGACGCTACACTGTCTGAACCACTGG	1853
141	lyGlyLysLeuLeuGlnGluTyrThrGlnLeuAla.....	152
1854	GTGGGAACCTTTGGAGAAATACCCAGCTTGCCAGGTATCAGGGAAT	1903
152	152
1904	AGGAGGAGAGAGTAGAATCTTGTGTGACAGTGCCTCGGATTGCCAGA	1953
152	152
1954	TATGAGGTAGAACCTTGAGAACTGACGACAGCTTGTCAGTGCCTTC	2003
153ArgGlyCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArg	167
2004	CCCCAGGGAATGTGACACTCTGGCTGTCTCTTGAGGTTCATATGAGCT	2053
168	GlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLe	184
2054	GGCCAAAGACTGGGAGCAGACTCAGAAATATCTCAATGTGTACGTCTCT	2103
184	uasSer..Lys.....	187
2104	GAACAGCAAAAGGTGAGACCTTATTAACCTTTAGCCTGCCTTCCCATGC	2153
187	187
2154	TCATTCTACCTAGATTCTCCAGATTGTTCACACTTATTTCTCTTAC	2203
187	187
2204	CCAAGGATTTAGGGGTGTGTCTACTTCACTTCCAGCTATTAACCTATCT	2253
187	187
2254	CTCTCTGGAGAGGTAAACCAAGGCTTCTAGAACAACCAAGCACTGATATTC	2303
188GlyAlaValValAlaThrTyrArgLysThrHisL	199
2304	CTTCTTTCTACTGTATGGGGCGAGTACTGTGCCTTACAGGAAGACACATC	2353
199	eucCysAspValGluIleProGlyGlnGlyPrometCysGluSerAsnSer	215
2354	TGTGTGACGTAGAGATTCCAGGGCAGGGGCTATGTGTGAAGAACAACCT	2403
216	ThmetProGlyProSerLeuGlnSerProValSerThrProAlaGlyLy	232
2404	ACCATGCTGTGGGCCAGACTTGTAGTCACTGTACACACACAGCAGGCGAA	2453
232	sileGly..Leu.....	235
2454	GGTAGAGGTTGTAAAGATGAGGGAAGGGGAACAGGAATACTTTGAACTG	2503
235	235
2504	GCAGTAGAGATAGAAAGCCCTTAAGAGAGGGGCTAATGGAATATAGACTA	2553
235	235
2554	GATGCTGTGTACAACAAGAGCAGGAAAGACTTAAGTAGGCTGTTTTTCAT	2603

```

236 .....AlaValCysTyrAspMetArgPheProGluLeuSer 247
      |||||
2604 TCAGATTTGGTCTAGCTGTCTGTATGACATGGGTTCCCTGACACTCT 2653
      |||||
248 LeuAlaLeuAlaGlnAlaGlyAlaGluLeuThrTyrProSerAla 264
      |||||
2654 CTGGCATTTGGCTCAGCTGAGCAGAGATTAATCTATCTTACGCTTT 2703
      |||||
264 egiysrliethrglyProAlaHisTrpGluVal..Leu..... 276
      |||||
2704 TGGATCATTTACAGGCCACGCCACCTGGAGGTAAAGATGATGCTTTT 2753
      |||||
276 ..... 276
2754 AATCATTAAGGGCTTTTCTTAACCTCATCTTCCCTTGCCCTTACCA 2803
      |||||
276 ..... 276
2804 TTAAATTCCTTCCCTTCCACCTAATGGGAAACATCATTCACAGAT 2853
      |||||
276 ..... 276
2854 TTCTCTCAGTAATAGTTAAATAGTAATCATCTTCTAGCAATTACCA 2903
      |||||
276 ..... 276
2904 ATAGTCACATAGGGTAGATTGCTGTAAATGTCCCTCACCTGTACCT 2953
      |||||
276 ..... 276
2954 CCACATTTTCTACATAGTACTTAAGTAGAACACATCTCATGCCAGG 3003
      |||||
277 ...LeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAla 292
      |||||
3004 TTGCTCGGGGCCGCTGCTATCGAACCAGTGTATGTAGTGGCAGCAG 3053
      |||||
292 aglncysgllyarghisglulysargalaserlyglyhissermetv 309
      |||||
3054 ACAGTGTGGAGCGCCACCATGAGAGAGAGCAATTTATGGCCACAGCAT 3103
      |||||
309 alyalasppttrpglythrvalvalalargcyssergluglyprogly 315
      |||||
3104 TGGTACAGCCCTGGGAGACAGTGTGGCCGCTGTGAGGGGCCAGGCG 3153
      |||||
326 LeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArg 342
      |||||
3154 CTCTGCTTGGCCGAAATAGACTCTGACTATCTGCGACAGTTGGCGACA 3203
      |||||
342 sLeuProValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGly 359
      |||||
3204 CCGGCTGTGTTCCAGCAGCGAGCGCTGACCTATGGAATCTGGGTCT 3253
      |||||
359 isProLeuSer 362
      |||||
3254 ACCCACTGTCT 3264
      |||||
seq_name: gb_to:AF069988
seq_documentation_block:
LOCUS AF069988 1338 bp mRNA linear ROD 23-JUL-1998
DEFINITION Mus musculus nitrilase 1 (Nlt1) mRNA, complete cds.
ACCESSION AF069988
VERSION AF069988.1 GI:3228667
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1338)
AUTHORS Bekarsky,Y., Campigillo,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Wernuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

```

```

TITLE Nitrilase and fhl1 homologs are encoded as fusion proteins in
JOURNAL drosophila melanogaster and caenorhabditis elegans
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
REFERENCE 98337986
AUTHORS 2 (bases 1 to 1338)
        Bekarsky,Y., Campigillo,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
        Tillib,S., Draganescu,A., Wernuth,P., Rothman,J., Huebner,K.,
        Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
FEATURES
source
    1..1338
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /chromosome="1"
        /map="1q21-q23"
        1..1338
        /gene="Nlt1"
        58..1029
        /gene="Nlt1"
        /codon_start=1
        /product="nitrilase 1"
        /protein_id="AAC40185.1"
        /db_xref="GI:3228668"
        /translation="MLGFTTRPPLQLCTGYRLTRPYLCTQPPRRTMSSTSWELPL
        VAVCOVSTPKROENFKALVOEAPLAEFLPEAFPIARNPAETLLSEPLN
        GDILGOYSOLAREGIMSLGFERHGDQWONKYNCHVLSKGSVVASVYRTHL
        CDVEIPROGPRRESNYTRPGSTLEPPYKTRGKGLATCYMRPELSIKTAQCAET
        LTPSARGSVTPAHWEYLARAIIESQVIAAQCGRHETASVGHSHVVDPMGT
        VVARCSSEPGICLARIDHFLQDMRHLVPHVRPRLDYSGLGHPLS"
BASE COUNT 347 a 335 c 330 g 326 t
ORIGIN
alignment_scores:
    Quality: 1496.00      length: 356
    Ratio: 4.704          Gaps: 3
    Percent Similarity: 89.326      Percent Identity: 80.337
alignment_block:
US-09-357-675c-25 x AF069988 ..
Align seg 1/1 to: AF069988 from: 1 to: 1338
      7 LeuAlaProAspArgProProAspArgThrLeuArgMetValLeuAlaI 23
        |||||
      2 CTGGCTCTGGCAGCTCTAGAGATCCGACCCGCTGATGTTTGGCTGT 51
        |||||
      23 eSerSerCysArgThrTyrSerLeuSerArgArgProArgLeuGlyPheI 40
        |||||
      52 GTCTTCATGCTG.....GGCTTCA 70
        |||||
      40 LeThrArgProProHisArgPheLeuSerLeuLeuGlyLeuArg 56
        |||||
      71 TCACCAAGGCTCTTCAACCA.....CTCCTGTATACCGGATCCGA 111
        |||||
      57 IleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAl 73
        |||||
      112 TTACTTGCAGACCCAGTACTTGTACTCAGCCAGGCGCCAGAACATGTC 161
        |||||
      73 alIeSerSerSerSerCysGluLeuProLeuValAlaValCysGlnVal 90
        |||||
      162 C...TCATCACTTCTCTGGAGAGCTGCCCTGGTGGCTGTGTGCAAGTAA 208
        |||||
      90 hsrSerThrProAspLysGlnGlnAsnPhelYsThrCysAlaGluLeuVal 106
        |||||
      209 CATCAACACCAACAACAAGCAAGCACTTTAAACATGTGCTGAGTTGGTT 258
        |||||
      107 ArgGluAlaAlaArgLeuGlnGlyAlaCysIleuAlaPheLeuProGlnAla 123
        |||||
      259 CAAGAGGCTGCCAGACTGGGTGCTTCTGCTTCTGCTGAGGCATT 308

```

```

123 eAspPheILeAlaArgSPRoAlaGluThrLeuHisLeuSerLuprol 140
|||||
309 TGACTTATATGCAGAAACCTGGCGAGACTATCTCTGTCGACACAC 358
|||||
140 euGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAlaArgGluCysGly 156
|||||
359 TGAAATGGGAGATCTTTGGCCAAATATAGCCAGCTTCCAGGAGATGTGGA 408
|||||
157 LeuThrLeuSerLeuGlyGlyPheHisGluArgGlyGlnLeuSPTpGlu 173
|||||
409 ATCTGGCTGTCTGGCGGTTTCCAGAGCGTGGCCAAAGCTGGAGCA 458
|||||
173 nThGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAla 190
|||||
459 GAATCAGAAATCTACAAATGTCTCATGTGCTTTTCAACAGCAGGATCAG 508
|||||
190 aValAlaThrTyrArgLysThrHisLeuCysAspValGluIleProGly 206
|||||
509 TAGTGGCCAGTTACAGAGACACATCTGTCCGATGTAGATCCACAGT 558
|||||
207 GlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerLeu 223
|||||
559 CAGGGCCGATGAGAGAAAGCACTATACCAAGCTGGAGGACCTTTGA 608
|||||
223 UserProValSerThrProAlaGlyLysIleGlyLeuAlaValCysTyr 240
|||||
609 GCCACCTGTCAAGACACCGCTGGCAAGTTGGTCTGACCAATCTGTATG 658
|||||
240 sPeKtarGpHeProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 256
|||||
659 ACATGGCGTTCCTCGAATCTTTGAAATGTGCTCAAGCTGGGCGAGAA 708
|||||
257 IleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHis 273
|||||
709 ATACTTACTATCTTCAAGCTTGTGATCTGTACAGTCCAGCCACTGT 758
|||||
273 pGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValAla 290
|||||
759 GGAAGTCTCTGTGGGGCCCGCCGATTTGAATCTCAGTGCATATTAATG 808
|||||
290 IaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHis 306
|||||
809 CAGCAGGCGAGTGTGAGCGCACCATGAACACAGCAAGTATGGCAT 858
|||||
307 SerMetValValAspProTyrPglTyrValAlaAlaArgCysSerGlu 323
|||||
859 AGCATGGTGTGACCGTGGGCGACAGTGTGCGCCGCTCCGAGCG 908
|||||
323 YProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeu 340
|||||
909 ACCAGGCTCTGCTGTCTGAATGTATCTTCACATTTCTACACAGATGC 958
|||||
340 rGArGHisLeuProValPheGlnHisArgArgProAspLeuTyrGlyAsn 356
|||||
959 GCCAACACCTGCGCTGTGTCAGACCGCAGACCTGACCTTATGGCAGT 1008
|||||
357 LeuGlyHisProLeuSer 362
|||||
1009 CTGGGTATCTCCATCTCT 1026
|||||
seq_name: gp_ro:BC021634
seq_documentation_block:
LOCUS BC021634 1365 bp mRNA linear ROD 22-JAN-2002
DEFINITION Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,
complete cds.
ACCESSION BC021634
VERSION BC021634.1 GI:18204912
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 1365)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT Contact: MGC help desk
Email: cgabds-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES
source
1..1365
/organism="Mus musculus"
/db_xref="LocusID:27045"
/db_xref="taxon:10090"
/clone="MGC:13825 IMAGE:4008543"
/tissue_type="Mammary tumor metastasized to lung.
MMTV-LTR/Int1 model. Expression driven by an MMTV-LTR
enhancer."
/clone_1lb="NCI-CGAP_Lu30"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
59..1030
/codon_start=1
/product="nitrilase 1"
/db_xref="GI:18204913"
/transcript="MLG17RPPEHOLCTGYRLRPVLTCTOPRPTMSSSTSMELPL
VAVCOVSTPNKOEENFKCALOEAGARAGACAPLEAFDFIARNAPATLILSEPL
GDLGOVSLARPCGIWLSLIGHERGOMDEONOKLYNCHVILNSGVASVSRKHL
CDEIETGOGPMRESNTYKPGTLEPVPKTPAGVGAICITDNFPELSLKLQAGMEI
LTVSSAFGSVTGPAAHWEVLLRARATSSQCVIAAOCGRHETRASVGHSMVVDPMGT
VVARCSGEGGLCTARIDLHFLDMDRHLDPVFOHRRPDLGSLGHPUS"

BASE COUNT 368 a 334 c 331 g 332 t
ORIGIN
alignment_scores:
Quality: 1478.00 Length: 349
Ratio: 4.707 Gaps: 3
Percent Similarity: 89.971 Percent Identity: 80.802
alignment_block:
US-09-357-675C-25 x BC021634 ..
Align seg 1/1 to: BC021634 from: 1 to: 1365
14 AspArgThrLeuArgMetValLeuAlaIleSerSerCysArgThrTyrSe 30
|||||
24 GATCGGACCCCGTGAATGTTGGCTGTGCTTATCTGTG..... 64
|||||
30 rLeuSerArgArgProArgLeuGlyPheIleThrArgProPheHisArgp 47
|||||
65 .....GGCTTCATCAGCAGGCTCTCTACCGAA. 91
|||||
47 heLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerValLeu 63
|||||
92 .....CTCTGTGTACCGGATACCGATATCTTCAACCCAGTACTT 133
|||||

```

```

64 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysG1 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 TGATCATGAGCCAGCCAGAACCATGTC...TCATCACTTCCTGGGA 180
80 uleuProLeuValAlaValCysGlnValThrSerThrProAspLysGln 97
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 GCTGCCCTGGTGGCTGTGTGCGAGGTAACATACACCAACAGCAG 230
97 lnaSnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 AGAAGCTTTAAACATGTGCTGAGTGTGTTCAAGAGCTGCAACAGGGT 280
114 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
281 GCTGCCCTGGCTTCTGCTGAGCATTTGACTTATTCACAGAAACCC 330
130 oAlaGluThrLeuHisLeuSerGluProLeuGlyLysLeuLeuGlnG 147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 TGCCGAGACATTACTCTCTGCGAACCCAGTAATGGGATCTTTGGGCC 380
147 lutyThrGlnLeuAlaArgGluCysGlyLeuThrPheSerLeuGlyGly 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
381 AATATGACCAAGCTTGCCAGGATGTGAATCTGGCTCTTGCGCGGT 430
164 PheHisGluArgGlyLysAspTrpGluGlnThrGlnLysIleTrpAsn 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
431 TTCACAGAGCGTGGCCAGACTGGAGACAGAAATCAAGAAATTAATG 480
180 shiValIleLeuAsnSerLysGlyAlaValAlaValAlaThrTrpArg 137
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
481 TCATGTGCTTTTGAACAGCAAGGATCAGTAGTGGCCAGTTACAGAGA 530
197 hciHisLeuCysAspValGluIleProGlyGlnGlyPrometCysGlu 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
531 CACATGTGCGCATGTAGAGATCCAGGTAGGGCCGATGACAGAAAGC 580
214 AsnSerThrMetProGlyProSerLeuGluSerProValSerThrPro 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
581 AACTATACAGAGCTGAGGACACTTTGAGCCACCTGTCAAGACACGCC 630
230 agLysIleGlyLeuAlaValCysTrpAspMetArgPheProGluLeu 247
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
631 TGCGAAGGTGGTGTAGCAATGTTATGACATGCGGTCCCGAACCTT 680
247 exLeuAlaLeuAlaGlnAlaGlnAlaGluIleLeuThrTrpProSerAl 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
681 CTTTGAATTTGGCTCAAGCTGGGGCAGAAATTACTTACTTATTTCTC 730
264 PheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAla 280
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
731 TTTTGATCTGTTAACAGTCCAGCCCACTGGAGAGTCTGCTGGGCGC 780
280 gAlaIleGluThrGlnCysTrpValAlaAlaAlaGlnCysGlyArgH 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
781 CGCCATTGAATCTCAGTGTATGTATAGCAGCAGCACTGTGGAGCC 830
297 lshiGluLysArgAlaSerTrpGlyHisSerMetValValAspProTr 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
831 ACCATTAACAGAGCAAGTATAGGCCATAGCATGTGTGACCCGTGG 880
314 GlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAla 330
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
881 GGCACAGAGTGGCCCGCTGCTCCGAGGACACAGCCCTGCTGCTGC 930
330 gIleAspLeuAsnTrpLeuArgGlnLeuArgGlnHisLeuProValPhe 347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
931 AATGTATCTCCACTTTTACACAGATGCGCCAAACACTGCTGTCTTC 980
347 lnhisArgArgProAspLeuTrpGlyAsnLeuGlyHisProLeuSer 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
981 AGCAGCGCAGAGCTGACCTCTATGCGCAGTGTGGTCAATCCACTCT 1027
seq_name: gb_htg:AC084821

```

```

seq_documentation_block:
LOCUS      AC084821                211771 bp    DNA       linear   HTG 29-JAN-2002
DEFINITION Mus musculus chromosome 1 clone rp23-395h6, WORKING DRAFT SEQUENCE.
ACCESSION  AC084821
VERSION    AC084821.24
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 211771)
AUTHORS    Deschamps, S., Gu, W. and Roe, B. A.
TITLE      Mus musculus BAC Clone rp23-395h6
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 211771)
AUTHORS    Deschamps, S., Gu, W. and Roe, B. A.
TITLE      Direct Submission
JOURNAL    Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT    On Jan 29, 2002 this sequence version replaced gi:16076995.
            ----- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            * 1 211771: contig of 211771 bp in length.
FEATURES             Location/Qualifiers
     source            1..211771
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /chromosome="1"
                        /clone="rp23-395h6"
                        /clone_1lb="RPCI Mouse BAC Library 23"
BASE COUNT      55622 a 48686 c 49207 g 58231 t      23 others
ORIGIN
alignment_scores:
    Quality: 1105.50      Length: 676
    Ratio: 3.893          Gaps: 8
    Percent Similarity: 42.012      Percent Identity: 37.722
alignment_block:
US-09-357-675c-25 x AC084821 ..
Align seg 1/1 to: AC084821 from: 1 to: 211771
64 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysG1 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
975 TGATCATGAGCCAGCCAGAACCATGTC...TCATCAACTTCCTGGGA 1021
80 uleuProLeuValAlaValCysGlnValThrSerThrProAspLysGln 97
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1022 GCTGCCCTGGTGGCTGTGTGCGAGGTAACATACACCAACAGCAG 1071
97 lnaSnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1072 AGAAGCTTTAAACATGTGCTGAGTGTGTTCAAGAGCTGCAACAGGGT 1121
114 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

1172	GCTGTCCTGGCCTTCTCTGCCTGAGGACATTGACTTAATGACAGAAACC	1171
130	oalaglunrhleuHisLeuSerGluProLeuglyLyLeuLeung	147
		:
1172	TGCCAGACATTACTCTGTCCGAACCACTGAATGGGATCTTTGGGCC	1221
147	IuTyThrInLeuala.....	152
:::::		:
1222	AATATGCGCAGCTTGCCAGGTACAAGGGTAGGGTTTGGGGGGGGGGGT	1271
152	152
1272	TGTGTGTAGAACCTCCGTGGACAGTCTTGATGCCACAGAGG	1321
153 ArgLucysgLYL	157
1322	GGGTAGATTGTGTGTGTGTGTCCCCCCCCCCCCCCGGAAGATGTGSA	1371
		:
157	eutRpleuSerLeuglyLyPheHsgLuAqglyGlnAspTrpGLuGn	173
:::::		:
1372	TCTGGCTGTCTTGGCGGCTTCCACGAGCGGGCCAAGACTGGGAGCAG	1421
174	ThngInLyILeTyAspCyshIsValLeuIeuAnsarLySgLYALA..	189
1422	AATCAAATAATCTACAAATTGTATGTGCTTTTGAACACAAAGGTGAGAC	1471
190	val.....	190
1472	GTCATFAGCCCTAGTTGGCTTCCACGGCTCTCTACCTTCTCTCAC	1521
190	190
1522	TCAGATCTTCAGCATTTCTTCTCAACTGTGGCCTTGGGGTTTTATGC	1571
190	190
1572	TTCTGTTCCAGCCTTACAAGTACTCTTTGGAGAGAGCTAACCTGGAC	1621
190	190
1622	TTGTAGAGTGTGTACTGAGATTTTCTCTCTCTCTCTCTCTCTCTCTC	1671
191 ValAlathrTYRA	195
1672	TCTCTCTCTCTCTCTCTCTCTCTCTCTAGATCAGTATGGCAGGTACA	1721
195	rgLyShrHisLeucysAspValGluLieProglyIngelyPrometCys	211
1722	GGAGGACACATCTGTGCGATGTAGAGATCCCAGGCGAGGGCCGATGGA	1771
212	GluSerAsnSerThMePrGlyProSerLeuGlnUserProValSerTh	228
		:
1772	GAAAGAACTATATACCAAGCCTGGAGGCACCTTGGAGCCACTGTCAAGAC	1821
228	rProAlagly.Lys.....	232
	:::	:
1822	ACCGCTGGCAAGTGGGAGTTTGGAAAGGAATGGACTATGTGAAA	1871
232	232
1872	GTGGTACACCCCCCATTTGGTCAGAGCTAATGAMGCACTCAGTGTCTG	1921
233 IlegLYL	235
1922	TTGCAATGGACAGGGAAGATGAGTGGGTGTTCATTTCAAGTTGCTC	1971
235	euaIaValcystTyraPmeTarGpheProGluLeuSerLeuaLaLeuala	251
		:
1972	TACCAATCTGTATATGACATGGCGTCCCGAACCTTCTTTGAATGTGCT	2021
252	GlnAlaGlyALaGluLeuThryrProSerAlaPheGlySerLierTh	268
		:
2022	CAGCTGGGGCAGAAATACTTACTTATCCCTTGAGCCTTGGATGTGTATC	2071

268	rolgtpoalahlstp.....	273
2072	aggtccagccaccactggaggttaagaaatgcccttcttaacataatagcc	2121
273	273
2122	cttccttaaccttactttttagtagtggaagcagagtgtacaaaccttaa	2171
273	273
2172	tccaaacaccttagtgatagccacaggttgcaaacacagactgccttgac	2221
273	273
2222	taaaaagaaaaaamaaatagaaagctggcagtgagggggtgcacacctttaa	2271
273	273
2272	tccacagacctcagagagcagaaagcaggtgatctctgagtttagagccag	2321
273	273
2322	ccgtgctcagagtggttccagagacagccagggctacacagaaaaac	2371
273	273
2372	ctgtctcaaaaacaaaaaagcaaccacaccttcaaaagtgggcgctag	2421
273	273
2422	tgtgtcagtcctctcattcagcactgtgagcaaaagcagagatctc	2471
273	273
2472	tgtgtgcttaagccagccagggcttacacagagaccttatctcaaaagt	2521
273	273
2522	acaaacacaaacactcctccttctgactgattttcttccctttcatct	2571
273	273
2572	gctgaggaattcatattcttagctagtgctgagtcattgtaaaagattag	2621
273	273
2622	ttatttaacttagttggttagtagaagattggccacagtcgcccgctt	2671
273	273
2672	cttgcacttgccctatttaaacctctgattgttggaagtcacagcacgt	2721
274glu..valleuleuarglaaargalaleglurhnglcystyr..va	288
2722	catctcagtgctgctgctggcgcccgccattgaaatcagatgctatgt	2771
288	valalalalalagincysgylarghishisulysargalasertyg	305
2772	aaagagagagagcagcatgtggagccaccactaaacaagagcaactttag	2821
305	lyhisseretvalvalaaproptrpgl..tth..valvalalaaargysser	321
2822	gccatagcatgtggttgaccccgtaggcacagtgctggcccgctgctcc	2871
322	glugllyproglyleucysleualaargileaspleuasntyrleuargl	338
2872	gaaggacacagggccttgccctgctgcaattgattccactttctacacaa	2921
338	nleuaargahisleuoprovalpheginhisargarproaspleuirtyg	355
2922	gargccccaacacctgcctgtgttcagacccgagacctgacctgtatg	2971

[illegible]

112 euGIyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArg 128
 |||||
 1766 TGGGTCCTGGCCCTGGCCCTTCTGCTGAGCATTTGACTTATTGACAGA 1615
 129 AspProAlaGluThrLeuHisLeuSerGluProLeuGlyLysLeuLe 145
 ::::|
 1816 AACCTGCCGAGACATTACTCTGTCGAAACCACTGAATGGGATCTTTT 1865
 145 uGIuGIuTyTrhGlnLeuAlaArg..Glu..... 154
 |::::|
 1866 GGGCAATATATAGCCAGCTTGCCAGTACAAGGTAGGGTTGGGGGGGG 1915
 154 154
 1916 GGGGTGTGTGTAGAACTCCCTGTTGAGCAGTGTCTGAAATGCCAC 1965
 154 154
 1966 AGAGGGGGTAGAGTTGNGTGTCGTGTCGTGCCCCCCCCCCCAGG 2015
 155 ...CysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAs 170
 |||||
 2016 GAATGTGGGATCTGGCTGTCTGGGGGTTTCCACGAGCGTGGCCAGA 2065
 170 pTrpGluGlnThrGlnLysIleTyraSncysHisValLeuLeuAsnSerL 187
 |||||
 2066 CTGGGACGAATACGAAATCTACATGTGTCTTGTGAACAGCA 2115
 187 ys..... 187
 2116 AGGGTGAGAGCTATAGCCCTTAGTTGCTCTTCCACAGCTTCTCTAC 2165
 187 187
 2166 CTTCCTCACTCAGATCTTACAGCATTTCTTCTCACTGCTGGCCTTG 2215
 187 187
 2216 GGGTTTATGCTTTTGTCTCAGCCTACAAAGTTACTCTTGGGAGAAAG 2265
 187 187
 2266 CTAAAGCTGAGCTGTAGAGTGTCTAGTACGTATTTCTTCTGTCTCT 2315
 188GlyAl 189
 2316 CT 2365
 189 aValValAlaThrTyrArgLysThrHisLeuGlyAspValGluLeuProG 206
 |||||
 2366 AGTAGTGGCCATTACGAGAAACACATCTGTGTGATGTAGATGCCAG 2415
 206 LysGlnGlyPrometCysGluSerAsnSerThrMetProGlyProSerLeu 222
 |||||
 2416 GTGAGGGGCCAATGAGAGAAAGCACTATATACCAAGCCTGGAGGACTCT 2465
 223 GluSerProValSerThrProAlaGly..Lys..... 232
 |||||
 2466 GAGCCACCTGTCAAGACACCGCTGGCAAGTGGAGTTGMAAAGGAA 2515
 232 232
 2516 TGGGACTATGTGMAAAGTGTAAACATCCCCCATTTGGTGCAAGCTAATGA 2565
 232 232
 2566 AGGCACTCACTGTTCTGTGATGGGACAGGAAAGATGAGTGGTGTGTTT 2615
 233IleGlyLeuAlaValCysTyrAspMetArgPheProGluL 246
 |||||
 2616 CTGATTTCAAGTTGTGTACCAATCTGTATGACATGCGGTTCCCTGAAC 2665

246 euserLeuAlaLeuAlaGlnAlaGlyValGluLeuThrTyProSer 262
 |||||
 2666 TTTCCTTGAATATGGCTCAAGCTGGGGCAGAAATACTTACTTACTTCA 2715
 263 AlaPheGlySerIleThrGlyProAlaHisTrp..... 273
 |||||
 2716 GCCTTGTGATGTGTACAGGTCCAGCCCACTGGGAGGTAGAGAAATGCT 2765
 273 273
 2766 CTTCATATACATATAGCCCTCTTAACTTACTTGTAGTGGAGCAG 2815
 273 273
 2816 AGTGATCAAACTAATATCCAGACTTCAAGTATAGACAGGTTCAA 2865
 273 273
 2866 AGCCAGCATGCTTGACTAAAAAAGAAACAAATAGAACCTGGGCAGTG 2915
 273 273
 2916 GGGGTGCACACTTTAATCCAGCACTCAGGAGGAGAGCAAGCAGGTGATC 2965
 273 273
 2966 TCTGAGTTGAGGCCAGCCTGCTGTACAGAGTGGGTCCAAGACAGCAG 3015
 273 273
 3016 GGCTACACAGAGAAACCTGTCTCAAAAACAAAAACCAACCCACACTT 3065
 273 273
 3066 CAAGAGTGGGGCATGTGTGTGATGCTCTTCAATCAGCAGCTGTGAGGC 3115
 273 273
 3116 AAAGCAGGCATATCTCTGTGGGTCTAGGCCAGCAGGTTACACAGGA 3165
 273 273
 3166 GACCTTATCTCAAAAGTAAACAAACACAAACACCTCTCTCTCTATGAT 3215
 273 273
 3216 TTTCCTCTTTTCATCTGTGAGGAAATCATTTCTTAGTAGGTGAG 3265
 273 273
 3266 TCATGTTAAGAGTTAGTTATTAACTAGTTGGTTAGTGAAGATGG 3315
 273 273
 3316 GCCCAGATGGCCGCTTCTTGTCACCTTGCCATTTAAAGCTGTATGTTCT 3365
 274Glu.ValLeuLeuAlaArgAlaIle 282
 3366 GGAATCAACGACGCTCTATCCCTCAGGTGCTGCGGGCCGCGCATTT 3415
 283 GluThrGlnCysTyrValValAlaAlaGlnCysGlyArgHisLysG 299
 |||||
 3416 GAATCTCAGTCTATGTATATAGCAGCGCAGTGTGAGCCCAACATGA 3465
 299 uLysArgAlaSerTyrgLysHisSerMetValValAspProTrpGlyThrV 316
 |||||
 3466 AACAAAGAGCAAGTTATGGCCATATGATGATGTGTTGACCCGAGGAGCAG 3515
 316 aValAlaIleArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAsp 332
 |||||
 3516 TGGTGGCCCGTGTCTCGAGGGACAGGCCCTTCTGCTTGTGCAATTGAT 3565
 333 LeuAsnTyLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArg 349

49539 TGGTAACACCCCCCATGTCAGAGCTAATGAAGCAGCTAGTGTCTGT 49588
 233IleGlyLe 235
 49589 TGAATGGACAGGAGATGAGTGGTGTCTTCATTTCAGGTGGCT 49638
 235 uAlaValCysTyrAspMetArgPheProGlyLeuSerLeuAlaLeuAlaG 252
 49639 AGCAATGCTGTATGACATGGCGTCCCTGAACCTTCTTGAATGGCTC 49688
 252 lAlaGlyAlaGlyIleLeuThrTyrProSerAlaPheGlySerIleThr 268
 49689 AAGTGGGGGAGAAATACTTACTTATCTTCAGCCTTGGATCTGTACA 49738
 269 GlyProIleHisTrp..... 273
 49739 GGTCCAGCCCACTGGGAGGTAGAGAAATGCCTTTCATACATATAGGCC 49788
 273 273
 49789 TCCTTTAACTTACTTTAGTATGGAGACAGAGTGGTACAACCTAAT 49838
 273 273
 49839 CCAAGCACTTCAGTGTAGTACCCACAGGTTCAAGCCAGCATGCTGACT 49888
 273 273
 49889 AAAAAAGAAAAAATAAGAAAGTGGCGCAGTGGGGGTGCACACCTTAAAT 49938
 273 273
 49939 CCCAGCACTCAGAGGAGCAGAAGAGTGATCTCTGAGTTTGAGGCCAGC 49988
 273 273
 49989 CTGGTCTACAGAGTGGGTTCACAGACAGCCAGGCTACACAGAGAAACC 50038
 273 273
 50039 TGTCTCAAAAAAACAACCAACCCACACTTCAAGAGTGGGCATGTGT 50088
 273 273
 50089 GGTCATGCTCTTCTATCTCAGCACTGTGAGCAAGGAGGAGAGATCTCT 50138
 273 273
 50139 GTGGGCTTAAGCCAGCAGGCTTACACAGGAGAGACCTTATCTCAAAAGTA 50188
 273 273
 50189 CAAACACAAACACTCTCTTCTACTGTGATTTCTTCTTTCATCTG 50238
 273 273
 50239 GTGAGGAATTCATTCTTACTAGTGTGTGATCTTAAAGAGTTAGT 50288
 273 273
 50289 TATTAACTCTAGTGTGTAGTGAAGATTTGGCCACAGTCCCGCTTC 50338
 273 273
 50339 TTGTCACTTGGCTTATTAACCGCTGTATGTTCTGAGTCACGACGTTCTC 50388
 274Glu.ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal 288
 50389 ATCTCTCAGGTCTGCTGGGGCCCGCCCATTTGAATCTCAAGTCTATGTA 50438
 289 ValAlaAlaAlaGlnCysGlyArgHisHisGlyIleValSerArgAlaSerTyrG 305
 50439 ATACGACAGCGCAGTGTGAGAGCCACCATGAAACAGACAGAGTTATGG 50488

305 YHisSerMetValValAspProTrpGlyThrValValAlaAlaArgCysSerG 322
 50489 CCATAGCATGTGGTGTGACCCGTCGGCAGAGTGTGGCCCGCTGCTCCG 50538
 322 lGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGln 338
 50539 AGGACCGAGGCTCTGCTGCTCGAATTTGATCTCATTCTACAAAG 50588
 339 LeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrG 355
 50589 ATGGCCCAACACCTGCTGTCTTCAGCAGCAGCAGCAGCAGCAGCTTARG 50638
 355 yAsnLeuGlyHisProLeuSer 362
 50639 CAGTCTGGCTATCCTCTCT 50660
 seq_name: gb_hhg:AC105589
 seq_documentation_block:
 LOCUS AC105589 178418 bp DNA linear HTG 09-JAN-2002
 DEFINITION Rattus norvegicus clone CH230-242B2, *** SEQUENCING IN PROGRESS
 *** 65 unordered pieces.
 ACCESSION AC105589
 VERSION AC105589.1 GI:18092812
 KEYWORDS HTG: HTGS PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 178418)
 REFERENCE
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alibrooks,S.L., AmaralLunge,H.C., Are,U.R., Banks,T., Barberia,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
 Carter,M., Cavazos,S.R., Checko,J., Chavez,D., Chen,R.,
 Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Geo,U., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgeson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homsli,F., Howard,S., Huber,J., Hulky,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyet,S.,
 Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,J., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtearge,O., Lieu,C., Liu,J., Liu,W.,
 Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,
 Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Plums,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshari,N.,
 Slisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Vellatton,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Wiczek,R., Wooden,S.,
 Watlington,S., Williams,G., Williamson,A., Wiczek,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.
 TITLE
 JOURNAL
 Direct Submission
 Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

2 (bases 1 to 178418)
Worley, K.C.
Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNGI

Center clone name: CH230-242B2

----- Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findphraplist

Consensus quality: 132588 bases at least Q40

Consensus quality: 142233 bases at least Q30

Consensus quality: 152072 bases at least Q20

Estimated insert size: 141845; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.8x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 8302: contig of 8302 bp in length
* 8303 8402: gap of unknown length
* 8403 18472: contig of 10070 bp in length
* 18473 18572: gap of unknown length
* 18573 24922: contig of 6350 bp in length
* 24923 25022: gap of unknown length
* 25023 30899: contig of 5877 bp in length
* 30900 30999: gap of unknown length
* 31000 38320: contig of 7321 bp in length
* 38321 38420: gap of unknown length
* 38421 43244: contig of 4824 bp in length
* 43245 48228: gap of unknown length
* 48229 48328: gap of unknown length
* 48329 54595: contig of 6267 bp in length
* 54596 54695: gap of unknown length
* 54696 58551: contig of 3856 bp in length
* 58552 58651: gap of unknown length
* 58652 62304: contig of 3553 bp in length
* 62305 62304: gap of unknown length
* 62305 66587: contig of 4283 bp in length
* 66588 70928: gap of unknown length
* 70929 71028: contig of 4241 bp in length
* 71029 75627: gap of unknown length
* 75628 75727: gap of unknown length
* 75728 78372: contig of 2545 bp in length
* 78373 78372: gap of unknown length
* 78373 80301: contig of 1929 bp in length
* 80302 80401: gap of unknown length
* 80402 83675: contig of 3274 bp in length
* 83676 83775: gap of unknown length
* 83776 87173: contig of 3398 bp in length
* 87174 87273: gap of unknown length
* 87274 90196: contig of 2923 bp in length
* 90197 90296: gap of unknown length
* 90297 92065: contig of 1769 bp in length
* 92066 92165: gap of unknown length
* 92166 94612: contig of 2447 bp in length

94613 94712: gap of unknown length
* 94713 94713 98755: contig of 403 bp in length
* 98756 98855: gap of unknown length
* 98856 102543: contig of 3688 bp in length
* 102544 102643: gap of unknown length
* 102644 105833: contig of 3190 bp in length
* 105834 105933: gap of unknown length
* 105934 108499: contig of 2566 bp in length
* 108500 108599: gap of unknown length
* 108600 111672: contig of 3073 bp in length
* 111673 111772: gap of unknown length
* 111773 114662: contig of 2890 bp in length
* 114663 114762: gap of unknown length
* 114763 116145: contig of 1383 bp in length
* 116146 116245: gap of unknown length
* 116246 118147: contig of 1902 bp in length
* 118148 118247: gap of unknown length
* 118248 120579: contig of 2332 bp in length
* 120580 120679: gap of unknown length
* 120680 122618: contig of 1939 bp in length
* 122619 122718: gap of unknown length
* 122719 124539: contig of 1821 bp in length
* 124540 124639: gap of unknown length
* 124640 126511: contig of 2012 bp in length
* 126512 126751: gap of unknown length
* 126752 128460: contig of 1709 bp in length
* 128461 128560: gap of unknown length
* 128561 130592: contig of 2032 bp in length
* 130593 130692: gap of unknown length
* 130693 132864: contig of 2172 bp in length
* 132865 132965: gap of unknown length
* 132966 134681: contig of 1917 bp in length
* 134682 134982: gap of unknown length
* 134983 136682: contig of 1701 bp in length
* 136683 136782: gap of unknown length
* 136783 139138: contig of 2356 bp in length
* 139139 139238: gap of unknown length
* 139239 140757: contig of 1519 bp in length
* 140758 140857: gap of unknown length
* 140858 142558: contig of 1701 bp in length
* 142559 142659: gap of unknown length
* 142660 144303: contig of 1645 bp in length
* 144304 144403: gap of unknown length
* 144404 145490: contig of 1087 bp in length
* 145491 145590: gap of unknown length
* 145591 147171: contig of 1581 bp in length
* 147172 147271: gap of unknown length
* 147272 148639: contig of 1368 bp in length
* 148640 148739: gap of unknown length
* 148740 150295: contig of 1556 bp in length
* 150296 150395: gap of unknown length
* 150396 151605: contig of 1110 bp in length
* 151606 151606: gap of unknown length
* 151607 152813: contig of 1208 bp in length
* 152814 152913: gap of unknown length
* 152914 155423: contig of 2510 bp in length
* 155424 155523: gap of unknown length
* 155524 156880: contig of 1357 bp in length
* 156881 156980: gap of unknown length
* 156981 158045: contig of 1065 bp in length
* 158046 158145: gap of unknown length
* 158146 159255: contig of 1110 bp in length
* 159256 159355: gap of unknown length
* 159356 160417: contig of 1062 bp in length
* 160418 160517: gap of unknown length
* 160519 161644: contig of 1127 bp in length
* 161645 161744: gap of unknown length
* 161745 163180: contig of 1436 bp in length
* 163181 163280: gap of unknown length
* 163281 164483: contig of 1203 bp in length
* 164484 164583: gap of unknown length
* 164584 166039: contig of 1456 bp in length
* 166040 166139: gap of unknown length

```

* 166140 167290: contig of 1151 bp in length
alignment_scores:
    Quality: 1064.00    Length: 836
    Ratio: 3.512        Gaps: 11
    Percent Similarity: 36.244    Percent Identity: 32.057
alignment_block:
US-09-357-675c-25 x AC105589 ..
Align seq 1/1 to: AC105589 from: 1 to: 178418

36 ArgLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 52
|||||
39092 AGGCTGGGCTTCATCACCAGGCTCTCATCACTCCCTGCTTCTGTG 39141
52 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro... 67
|-|-|
39142 TACCGGATACCGGATTTACCTCAATCTCAGTACTTGTACTCAGCCAGGT 39191
67
39192 AACATATTTATGTTCTTGTCTAGACAGTAGTGTCAAGTCACTGTTTA 39241
67
39242 TTACCCGAATAGTGACCTGCCATCCACCAACATTTATGCGTGTACTC 39291
67
39292 CAGTAGCTTAGGCTTATTTAATGCTGTGAGAAATGAAACAGTGGCTA 39341
67
39342 GGGTTGGTGAGACACACCTTAATCCAGACCTGGGAAACAGCAGAGAG 39391
67
39392 GATCACTAGAGATTAAATCATCCCGCTACATAGAAAGTCTGCTG 39441
67
39442 GGTACAAAAAAGACCGGTCTCAAAAGGGAGGTAAATCAGAAAAAAA 39491
67
39492 AAGTAGTTCCGGCTCCACCCCACTTGAGCTCCCAAGTTTCTCCGG 39541
68
39542 GCACGCGGTAGAAATCCTGAGAAACCTGCTGTGCATTCACAGGCCAG 39591
70
39592 AGGCATGGCAGAGTCTCATGACTTCTCGGAGCTGCCCTGGTGGCTG 39641
86
39642 TGTCGCCAGGTAACTCAACCAACCAAGCAAGAACTTTAAACGCTGT 39691
103
39692 GCTGAGCTGTTCAAGAGGCTACAGAGCTGGCGCTTGCCTGGCTTCT 39741
119
39742 GCCTGAGGCAATTTGATTATTCAGCAAAATCCTGCCAGACATTTACCC 39791
136
39792 TGTCCGAACCACTGATGGGAGCTTTTGGAGCAATATATAGCCAGCTTCC 39841
152
152

```

```

39842 AGTACAAAGGCTAGGGGTGGGGGAGGGGTANANATCCCTGTGGG 39891
152
39892 CAGTGTCCCTGGATTGCCACATAGAAAGGGGTAGAGTGTGCATGCTCC 39941
153
39942 CCCCTCAGGAATGTGAATCTGGCTGTCTGGGTGGTTCACCAACAG 39991
167
39992 TGGCCCAAAATGGGAGACAGAAATCCGAAATCTACCATTTGCTGGGCTC 40041
184
40042 TGAACAAACCAAGGTGAGACATTCCAATCTCAACTGGCTTCTCTAG 40091
187
40092 GTCACTACCTTCTCTGCTCAGTAGAATTTAACAATTTTTCACCTGCTG 40141
187
40142 GCTTAGGGGTTTTTTTGTGTTTGAATGCTTCTGTCTTACCTAA 40191
187
40192 AAATTACTCTTTGGGAGAAGCTAAGCTGACCTCCAGTGTCTTACTG 40241
188
40242 AGCATTTTCTCTTACTTATGATCAGTGTGCGCAGTTACAGAGAGAC 40291
198
40292 ATCTATGTGATGTAGAGATCCAGGTCAGGGCCATATGAGAAAGCAAC 40341
215
40342 TATACCATGCTCGATACGCTCTTGAACCACTGTCAAGACACAGCAGG 40391
231
40392 CAAGGTAGAGATTGAAAGAAATAGAAATATGTGAAAAAGTTGTAAC 40441
235
40442 ATCTCCCTTGTGTAGATGATGAGGTAGTATTCATGTACTGTGAC 40491
236
40492 CAGGACAGGAAGATGAGTGGTGTCTTCTGCTTCAGAGGTGCTTAGA 40541
237
40542 ATCTGTATGATGATGCGGTCCCTGACCTTCTGTAATGAGGCTCAAGC 40591
253
40592 TGGGCCAATAATCTTACTTATCTTCAAGCTTTTGATCTGTACCGGCC 40641
270
40642 CAGCCCACTGGAGGTAGAGAAATCCTCTTCATACATATAGGCTCTCT 40691
275
40692 TTAATCTATTTTAGTAGTGGGAGAGATAGACAAAACCTAAACTCTA 40741
275
275

```

```

40742 AGCACTTCAGGGGATAGCCACAGGTTCAAAAGCCAGCATAGTATGATGCTCG 40791
275 .....
40792 AGTAAAAACAGAAAAAATAGAAAGCTGGGACAGTGGGGGTGCACACCTTT 40841
275 .....
40842 AATCCAGCCCTCAGAGGAGCAAGAGCAAGTGTGATCTGATTTCAAGACCAG 40891
275 .....
40892 CCGTGCTACAGAGTAGTTCCAGAGCAGCAGGGCTTACACAGAGAAACC 40941
275 .....
40942 GTCTCAAAAACAAAACAAAACAAAACAAAAGTAAACCACACACC 40991
275 .....
40992 AAAATAGTGTGATGCTGGTGGTGGCTTTTATCTCCGACATGTAAAGC 41041
275 .....
41042 AGAGCCCGCAGATCTGTGGGTCCAGGCCAACAAAGTTATACAAAGA 41091
275 .....
41092 GACCTTATATAAAAAAATACAAAACAAACACACCCCTCCCTTCTACT 41141
275 .....
41142 TGAAGATTTTCTCCCTCCGATCTGGTGAGAAATTCGTTCTTAGAT 41191
275 .....
41192 AGGTGTGATCATGTTATAGTACTAGTAACTTCTAGTTCATATAGTC 41241
275 .....
41242 AATGATGATTTGGGCCACAGTCCGCTTCTGTCTACCTCCCTATTAA 41291
276 .....
41292 ATGCTGATGTCTCTCAGTACGACGCTCATCTCCTCAGTGTGCTGCGG 41341
279 AAlaArgAlaIleGluThrGlnCysTyrValAlaAlaIleGlnCysG 295
41342 GCCCGGCGCATTTAGTCTCAGTCTAGTATAGCAGCAGCAGCAGTGTGG 41391
295 YATGHTSHISGILuLysArgAlaSerTyrGlnHisSerMetValAlaAsp 312
41392 ACCGACCAATGAGACAGACAGAGTTACGGCCACAGCATGTGGTTTACC 41441
312 GTPRGLIYHrValAlaIleArgCysSerGluLysProGluLysLeu 328
41442 CCGTGGGACAGTGTGGCCAGCTGTCCGAGGGACAGGCTTTGCTT 41491
329 AAlaArgIleAspLeuAsnTyrIleuArgGlnIleuArgHisIleuProVa 345
41492 GCTCGAATTTATCTCCACTTTGTGCACAGATGGCCACACTTGCTGTG 41541
345 lPheGlnHisArgArgProAspLeuTyrGlnLysLeuGlnHisProLeus 362
41542 GTTCCAGCATCGCAGACCTTATATGCGACAGTCTC...CTGCCACTCT 41588
362 er 362
11
41589 CC 41590
seq_name: gp_ov:AF284575
seq_documentation_block:

```

```

LOCUS AF284575 1214 bp mRNA linear VRT 22-JUL-2000
DEFINITION Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.
ACCESSION AF284575
VERSION AF284575.1 GI:9367117
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 1214)
Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
Bleganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
Crystal structure of the worm Nit1hit Rosetta Stone protein reveals
a Nit tetramer binding two Phit dimers
Curr. Biol. 10 (15), 907-917 (2000)
JOURNAL MEDLINE
20414396
REFERENCE
2 (bases 1 to 1214)
Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,
Bleganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.
Direct Submission
Submitted (05-JUL-2000) Kimmel Cancer Center, Thomas Jefferson
University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA
FEATURES
source
1..1214
/organism="Xenopus laevis"
/db_xref="taxon:8355"
1..1214
/gene="Nit1"
180..1046
/gene="Nit1"
/note="Nit1: similar to the Nit domains of Caenorhabditis
elegans Nit1hit and Drosophila melanogaster Nit1hit"
/codon_start=1
/product="Nit protein 1"
/protein_id="AF87104.1"
/db_xref="GI:9367118"
/translation="MAGAHKPLIAYCOMTSTSDKEKNFATCSRLIREAGRRACWFL
PEAFYIGSIEETLSLAESLHGDTIQRYTOLARCGMLSLGFGHEKGPMDPDR
NSHVYVNTGTHIVSVYRAHLEFDVLDNGVLRSSSTLPGAEILRITSPAGKIG
GVCYDLRPEPSIALAOGCAELTYPPSAFTTGTGAHEVILRRARITGCVYVAAO
TDRAHEKRTSYGHANVYDPKGLVYIGCGEGTGTCAETLIDIFYMERVRDMVWRHRT
DLYKISFNKPD"
BASE COUNT 281 a 305 c 345 g 282 t 1 others
ORIGIN
alignment_scores:
Quality: 968.50 Length: 307
Ratio: 3.905 Gaps: 1
Percent Similarity: 80.782 Percent Identity: 57.980
alignment_block:
US-09-357-675C-25 x AF284575 ..
Align seg 1/1 to: AF284575 from: 1 to: 1214
51 LeuCySProGluLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67
|||||
144 CTGTGTCGGATATAGGGGTTAGAGCCTCAGTGCATG..... 182
67 oArgProArgAlaMetaIleSerSerSerSerCysGluLeuProLeuV 84
183 .....
84 alaIaValCysGlnValThSerThProAspLysGlnGlnAsnPhelys 100
:|||||
205 TTGCGGTGCCAGATGACTTCAACCTTCATAGGAGAAATTTTCGCC 254
101 ThnCysAlaGluLeuValArgGluAlaIleArgLeuGluValaCysLeuAl 117
|||||
255 ACGTGTTCGGGCTGATCGGGAGGCTGGCGGCGTCCGCTTGCATGTT 304

```

```

117  aPhelProgluAlaPheaspPheileAlaArgAspProAlaGluThrL 134
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305  GTTCTGCGGAGAGCTTGTGACTATATCGGGGACCATTTGAGGAGAGCC 354
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134  eunHisLeuSerGluProLeuGlyLysLeuLeuGluGluThrLysIn 150
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355  TGAGCTGTGCTGAGTCTCATATGAGGAGACCATTCACGCTTACCCCA 404
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151  LeuAlaArgGlyLysGlyLeuThrLeuSerLeuGlyLysPheHisGlu 167
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405  CTCGCCAGAGGTGGGTGGCTGCTGCTTCCCTGGGGGATTTCAAGAGA 454
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167  gGlyAlaSerProGluGluThrGlnLysIleTyrAsnCyHisValLeuL 184
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455  AGGACCCCACTGGGACAGGACGACCAACGATTTCCATTTCTCACGGT 504
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184  euAsnSerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCyS 200
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
505  TGGACACACAGGCGCATAGTATCGTGTACCCGACAGGCTCACCTGTT 554
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201  AspValGluIleProGlyGlnGlyPrometycSgiUserAsnSerThm 217
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
555  GACGTAGACTTGCAGAAATGAGTGTACTCAGAGAGCAGTCCACCT 604
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217  tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIle 234
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605  CCCCAGACAGAGCTTATTCGCCCATCATCTTCTCACAGAGAGAAAT 654
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234  LLeuAlaValAlaCysTyrAspMetArgPheProGluLeuSerLeuAl 250
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
655  GCTGGGGGTGTGTACGACCTCCGCTCCAGAAATCTCTGCTGTG 704
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251  AlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySer 267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
705  GCCCAACAGAGGACGAACCTCTCTCACTTACCCTTCTGCTTACCC 754
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267  eThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleG 284
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755  TACTGTGTGGCAGCTTGGAGGTGTCTGAGAGCCGCTGCCATAGAAA 804
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
284  hArgLysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGlu 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
805  CCCAGTGTACGTAGTTCAGCGGACAGACAGACAGACAAATGAGAAG 854
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301  ArgAlaSerTyrGlyHisSerMetValAlaSprProTyrGlyThrVal 317
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
855  AGGAGCTCTTATGTGTACGCTATGTGTGTAGCCGCTGGGGCTGTGAT 904
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
317  LAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLe 334
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
905  TGGCCATGTCCAGAGGAGACAGGAAATGTATTATGCTGAGATTGACAT 954
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
334  snTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArg 350
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
955  CCTACATGAGAGCGTGTAGGCGGACATGCCGCTGGAGGACGACGAG 1004
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351  ProAspLeuTyrGlyAsnLeu 357
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1005  ACTGATCTGTATGGGAAATC 1025
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: gb_htg:AC017755
seq_documentation_block:
LOCUS      AC017755              131853 bp      DNA      linear      HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, ***, SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017755
VERSION     AC017755.1  GI:6553435
KEYWORDS    HTG: HTGS_PHASE2.
SOURCE      Fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

```

REFERENCE   1 (bases 1 to 131853)
AUTHORS    Adams, M. and Venter, J.C.
TITLE      Direct Submission
JOURNAL    Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDW:10211975 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.

FEATURES
    source
        1..131853
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
BASE COUNT   36037 a 28741 c 28528 g 38547 t
ORIGIN
alignment_scores:
    Quality: 715.50      Length: 364
    Ratio: 2.944         Gaps: 6
    Percent Similarity: 66.758      Percent Identity: 42.857

alignment_block:
US-09-357-675C-25 x AC017755 ..

Align seg 1/1 to: AC017755 from: 1 to: 131853

17  LeuArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSer 33
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84367  CTTCCTTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 84416
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
33  gArgPro..... 35
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84417  CCGTAAGCTAGGGGAATGTAAGAGAGCAGCGCATTAACAATAGTA 84466
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
36  ..... 42
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84467  AAAATTAATTGATTAACATCTTACGAGCTAACGATTCGMAAATTTCCACAG 84516
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43  ProProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIlePro 59
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84517  CCGGCGACCCGTTTCAATGTCACTGTACTGTAAATACCACGAGA..... 84560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59  InLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSer 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84561  GCAGCATTTGTATCCCATTCATCAGCAGCTACGCGAGATGTGCCGCA 84610
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76  SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSer 92
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84611  AAGAGAAAGATCAAAAGTCCGACCAATTCCTGTGGGTCAAGATCGCTTAC 84660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92  rProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGlu 109
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84661  CAGCGACAGCGCGGCTAATCTTAGCCAACTGATPAGAGCTAGTGATAGG 84710
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109  LAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPhe 125
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84711  CCAAGTACAGAAAGCGCTGATGCTCTTCTGCTGAGTGTGATGCTTT 84760
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
126  IleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGly 142
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84761  GTGGGCGAAGCGCGGACCCCAACTATTGAGCTCTCGAGGCGCTTGACGG 84810
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142  YLysLeuLeuGluGluThrThrGlnLeuAlaArgGlyLysGlyLeuThr 159
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84811  CGAGTTAATGGCGCAATACCGGGAATGGCAAGGCAACAAATTTGGA 84860
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159  euSerLeuGlyLysPheHisGluArgGlyLysAspThrPheGluThrGln 175
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84861  TTTCCTGTGGTGTGCGTGCAGAGCGGAAAGAT.....CNA 84895
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

176 LysIleTyrAsnCysHisValIleuLeuAsnSerLysGlyAlaValAlaI 192
      |||||.....:|||||.....:|||||.....:|||||.....:
84896 AAAATCTTCAACGCTCATGTTTCTCAACAGAGAAAGGCAATCAGACG 84945
      |||||.....:|||||.....:|||||.....:|||||.....:
192 aATHTYATGLYThrHisLeuCysAspValGluIleProGlyGlnGlyP 209
      |||||.....:|||||.....:|||||.....:|||||.....:
84946 AGTATACAGAAACCTGCACATGTTTGTGACACTTAA...GAGGTTCC 84992
      |||||.....:|||||.....:|||||.....:|||||.....:
209 roMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 225
      |||||.....:|||||.....:|||||.....:|||||.....:
84993 GCCTACGCGCATCAGATACAGTTACGCCGGGATCTCCTTGAGCCGCCA 85042
      |||||.....:|||||.....:|||||.....:|||||.....:
226 ValSerThrProAlaGlyLysIleGlyAlaValAlaCysTyrAspMetAr 242
      |||||.....:|||||.....:|||||.....:|||||.....:
85043 GTGAGCAGCTCAGATTGGCCAGATAGGCTTCAGATTGCTACAGACCTGC 85092
      |||||.....:|||||.....:|||||.....:|||||.....:
242 gPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuT 259
      |||||.....:|||||.....:|||||.....:|||||.....:
85093 TTTTGTGACGCGCGCTGCTCAGAGAAAGCTGGGCTGCCAATTGTTAA 85142
      |||||.....:|||||.....:|||||.....:|||||.....:
259 hTyrProSerAlaPheGlySerIleThrGlyProAlaHisTyrPgluVal 275
      |||||.....:|||||.....:|||||.....:|||||.....:
85143 CATACCATCCGATTCATACATACGCAACCGGTAAAGCGCACTGGGAATC 85192
      |||||.....:|||||.....:|||||.....:|||||.....:
276 LeuLeuArgAlaAlaIleGluThrGlnCysTyrValValAlaAlaAlaI 292
      |||||.....:|||||.....:|||||.....:|||||.....:
85193 CTCCTGGGGCCAGAGCCATAGACATCATCTTGTGCTGGCTGGCGC 85242
      |||||.....:|||||.....:|||||.....:|||||.....:
292 aGlnCysGlyArgHisIleGluLysArgAlaSerTyrGlyHisSerMetV 309
      |||||.....:|||||.....:|||||.....:|||||.....:
85243 TCAGATAGTTGGACACAAACGACAGAGATTGGGCGCACAGCATGA 85292
      |||||.....:|||||.....:|||||.....:|||||.....:
309 aValAspProTyrPgluThrValValAlaArgCysSerGluGlyPro... 324
      |||||.....:|||||.....:|||||.....:|||||.....:
85293 TCGTATGCCCTCGGGAACGTTACTGCTACTGACGCGAGAGAGACTT 85342
      |||||.....:|||||.....:|||||.....:|||||.....:
325 GlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgIleuArgAr 341
      |||||.....:|||||.....:|||||.....:|||||.....:
85343 GAAATAGGACGCGCGAGTGAGACTTCCGCTTCATCTCGTATCA 85392
      |||||.....:|||||.....:|||||.....:|||||.....:
341 gHisLeuProValPheGlnHisArgArgProAspLeuTyr 354
      |||||.....:|||||.....:|||||.....:|||||.....:
85393 GAACTATCCCTGCTCGACATCTGTCGAAACGACATCTAC 85432
      |||||.....:|||||.....:|||||.....:|||||.....:
seq_name: gbl_in:AC093121
seq_documentation_block:
LOCUS AC093121 168250 bp DNA linear INV 11-AUG-2001
DEFINITION Drosophila melanogaster 3L BAC RP98-11J11 (Roswell Park Cancer
Institute Drosophila BAC library) complete sequence.
ACCESSION AC093121
VERSION AC093121.1 GI:15148095
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 168250)
Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Guyanayre,J.D., Taber,P., Williamson,A., Homs,F.H.,
Dugan-Hochia,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Aylee,M.A., Scott,G.S., Worley,K.W., Amaratides,P.G., Brando,R.C.,
Rogers,Y., Au,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Daveenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Drapper,H., Emery-Cohen,A., Ferriera,S., Gary,N.D.S., Houck,J.,
Hoslin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,M., Mettel,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,Y.,
Schaefer,E., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

Wheeler,D., Weinstein,G., Gibbs,R. and Venter,J.C.
Direct Submission
Unpublished
2 (bases 1 to 168250)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Usman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blinaga,K., Blankenburg,K., Bonnin,D., Bouch,J.,
Bowle,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dethorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Hollway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,X., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kralovic,D., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisegeed,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,D., Newton,N.,
Nguyen,M., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogun,M., Okunoye,G., Oregunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Plums,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoostatli,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Submitted (11-AUG-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2

OM of: US-09-357-675c-25 to: N.Geneseq_032802:* out_format: pfs
Date: Apr 29, 2002 10:22 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-O=/cgn2.1/USFTO.spool/US09357675/runat.29042002_091757.3254/app-query.fasta.1.1208  
-MODE=frame+ p2n.model -DEV=yih  
-Db=N.Geneseq_032802 -QW=fastap -SUFFIX=p2n.rng -GAOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MAPFIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675.@CGN1_1.412 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THEADS=1
```

Search information block:

Query: US-09-357-675c-25
Query length: 362
Database: N.Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 338.150000

score list:

Sequence	Strd Orig	ZScore	EScore	len	! Documentation	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			1919.00	2947.61	7.8e-15	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			1806.00	2773.24	4.0e-14	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			1067.00	1632.70	1.4e-82	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			968.50	1480.40	4.1e-74	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			715.50	1078.53	1.0e-5	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			715.50	1073.71	1.9e-5	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			710.50	1079.60	8.7e-5	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			495.00	750.69	1.8e-3	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			485.50	732.89	1.8e-32	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			481.00	726.45	4.1e-32	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			404.00	566.81	3.2e-23	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			404.00	550.40	2.6e-22	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			404.00	536.01	1.7e-21	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			392.00	585.98	2.7e-24	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			387.00	584.90	3.1e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			386.50	572.01	1.6e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			385.00	567.75	2.8e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			301.00	436.24	4.6e-17	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			292.50	378.11	1.0e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			291.00	436.78	5.6e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			291.00	436.78	5.6e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			283.50	426.17	2.2e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			283.50	426.17	2.2e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			276.50	406.11	2.8e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			276.50	405.44	3.1e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			270.00	401.45	5.1e-14	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			270.00	401.46	5.2e-14	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			262.00	363.15	5.2e-12	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			248.00	351.51	3.1e-11	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			229.00	346.86	5.7e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			220.50	308.45	7.8e-9	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			218.00	325.00	9.4e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			218.00	309.58	6.8e-9	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			217.50	311.35	5.4e-9	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			213.50	309.80	6.6e-9	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			212.50	312.99	4.4e-9	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101 +			212.50	308.26	8.0e-9	

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AA056972 + 212.50 308.26 8.0e
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AA056993 + 212.50 308.26 8.0e
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AA056994 + 212.50 308.26 8.0e
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AA056973 + 211.50 306.71 9.8e
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:AA056968 + 211.50 306.71 9.8e

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ46101

seq_documentation_block:

ID AAZ46101 standard; DNA: 1416 BP.

AC AAZ46101;

DT 05-MAY-2000 (first entry)

XX Nucleotide sequence of the coding region of NIT1 gene.

DE NIT1 gene; nitrlase; tumour suppressor gene; FHIT; chromosome 3p14.2;

KW FRA3B; cancer; genome allele inactivation; ss.

XX Homo sapiens.

OS Mus sp. Drosophila melanogaster.

OS Caenorhabditis elegans.

FH Key Location/Qualifiers

FT CDS 3..1415

FT /*tag= a

FT /product= (pos: 18..20, aa: Xaa)

FT /product= (pos: 1179..1181, aa: Gln)

FT /product= (pos: 1182..1184, aa: Ala)

FT /note= "contains 7 internal stop codons; Xaa is an unspecified amino acid"

PN WO200003685-A2.

XX 27-JAN-2000.

PD 20-JUL-1999; 99WO-US16366.

PR 20-JUL-1998; 98US-0093350.

XX (UYJE) UNIT JEFFERSON THOMAS.

PI Croce CM;

XX WPI; 2000-171195/15.

DR P-PSDB; AAY68739.

XX Novel nitrlase homologs used as diagnostic and therapeutic reagents for the detection and treatment of cancer -

PT Claim 6; Fig 6; 25pp; English.

XX The present sequence represents the coding region of human, murine, Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The

CC human and mouse NIT1 genes are members of an uncharacterised mammalian gene family with homology to bacterial and plant nitrlases.

CC The tumour suppressor gene FHIT in D. melanogaster and C. elegans code for fusion proteins in which the Phit domain is fused with a Nit domain.

CC In mouse and humans, FHIT and NIT are encoded by two different genes, localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The

CC human FHIT gene at chromosome 3p14.2, spanning the constitutive chromosomal fragile site FRA3B, is often altered in most common forms

CC of human cancer. The NIT1 protein overcomes the mutated inactivation of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives

CC and analogues of them, and antibodies are used as diagnostic and therapeutic reagents for the detection and treatment of cancers.

CC

XX Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;

SQ

alignment_scores:

Quality: 1919.00 Length: 362
 Ratio: 5.316 Gaps: 0
 Percent Similarity: 99.724 Percent Identity: 99.724

alignment block:

US-09-357-675c-25 x AA246101 ..

Align seg 1/1 to: AA246101 from: 1 to: 1416

```

1  ProLeuAlaAlaAla**LeuAlaProAspArgProProAspArgThrLe 17
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
3  CCACCTGCGCGCCCTNTCTGGCTCCAGACGCGCCCTCGGATCGGACCT 52
   | | | | | | | | | | | | | | | | | | | | | | | | | |
17 uArgMetValLeuAlaAlaSerSerCysArgThrTyrrSerLeuSerArg 34
   | | | | | | | | | | | | | | | | | | | | | | | | | |
53 GCGAATGGTTTGGCTATATCTTTCATGTAGAGACTACTCCATCCCGC 102
   | | | | | | | | | | | | | | | | | | | | | | | | | |
34 rGProArGLeuGlyPheLeIleThrArgProProHisArgPheLeuSerLeu 50
   | | | | | | | | | | | | | | | | | | | | | | | | | |
103 GCGCGCGGCTGGGCTTCATCACACGAGCCCTCCACAGATTCCTGTCCTT 152
   | | | | | | | | | | | | | | | | | | | | | | | | | |
51 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnI 67
   | | | | | | | | | | | | | | | | | | | | | | | | | |
153 CTGTCTCCGAGACTCGGATACCTCACTCACTGACTTGTGCTCAGCC 202
   | | | | | | | | | | | | | | | | | | | | | | | | | |
67 oArgProArgAlaMetAlaAlaSerSerSerSerCysGlnLeuProLeu 84
   | | | | | | | | | | | | | | | | | | | | | | | | | |
203 CAGGGCCAGAGCCATGGCTATCTCTCTCTCTCTCTCGGAACTGCCCT 252
   | | | | | | | | | | | | | | | | | | | | | | | | | |
84 aAlaValAlcGlnValThrSerThrProAspLysGlnIAsnPhelys 100
   | | | | | | | | | | | | | | | | | | | | | | | | | |
253 TGGCTGTGTGCCAGGTACATCGACGCGCAGACAGACAGAACTTTAA 302
   | | | | | | | | | | | | | | | | | | | | | | | | | |
101 ThrCysAlaGlnLeuValArgGlnAlaArgLeuGlyAlaCysLeuAl 117
   | | | | | | | | | | | | | | | | | | | | | | | | | |
303 ACATGTCTGAGCTGTGAGAGGCTGCCAGACTGGGCTGCTGCTGCG 352
   | | | | | | | | | | | | | | | | | | | | | | | | | |
117 aPheLeuProGlnAlaAlaPheAspPheIleAlaArgAspProAlaGluThrL 134
   | | | | | | | | | | | | | | | | | | | | | | | | | |
353 TTTCTGCTGAGGCAATTGACTTCATTCAGCGGAGCCCTGACAGAGAGCC 402
   | | | | | | | | | | | | | | | | | | | | | | | | | |
134 euHisLeuSerGlnProLeuGlyGlyLysLeuLeuGlnIuTyrrThrGln 150
   | | | | | | | | | | | | | | | | | | | | | | | | | |
403 TACACCTCTCTAACCCACTGGGTGGAAACTTTTGGAAATAACACCCAG 452
   | | | | | | | | | | | | | | | | | | | | | | | | | |
151 LeuAlaArgGlnCysGlyLeuThrPleuSerLeuGlyLysPheHisGln 167
   | | | | | | | | | | | | | | | | | | | | | | | | | |
453 CTTCGACAGGAATGTGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCG 502
   | | | | | | | | | | | | | | | | | | | | | | | | | |
167 gGlyGlnAspTrpGlnGlnThrGlnLysIleTyrrAsnGlyHisValLeu 184
   | | | | | | | | | | | | | | | | | | | | | | | | | |
503 TGGCCAAAGCTGGAGAGCACTCAGAAATCTAACATTTGTCACGTGCTGC 552
   | | | | | | | | | | | | | | | | | | | | | | | | | |
184 euAsnSerLysGlyAlaValAlaAlaThrTyrrArgLysThrHisLeuGly 200
   | | | | | | | | | | | | | | | | | | | | | | | | | |
553 TGAACAGCAAAAGGGGAGTAGTGGCCACTTACAGAGAACACATCTGTCT 602
   | | | | | | | | | | | | | | | | | | | | | | | | | |
201 AspValGlnIleProGlyGlnGlyPrometCysGlnSerAsnSerThr 217
   | | | | | | | | | | | | | | | | | | | | | | | | | |
603 GACGTAGAGATTCAGGGGAGGGGCGCTATGTGTGAAGCAACTTACCAT 652
   | | | | | | | | | | | | | | | | | | | | | | | | | |
217 tProGlyProSerLeuGlnSerProValSerThrProAlaGlyLysIle 234
   | | | | | | | | | | | | | | | | | | | | | | | | | |
653 GCGTGGCGCCAGTCTTGAAGTCACTGTGACACACAGAGAGCAAGATTGG 702
   | | | | | | | | | | | | | | | | | | | | | | | | | |
234 lLeuAlaValAlcYsTyrrAspMetArgPheProGlnLeuSerLeuAla 250
   | | | | | | | | | | | | | | | | | | | | | | | | | |
703 GTCTAGCTGTCTGTATGACATGCGGTCCCTGGAACCTCTCTGTCGATGG 752
   | | | | | | | | | | | | | | | | | | | | | | | | | |
251 AlaglnAlaGlyAlaGlnIleuThrTyrrProSerAlaPheGlySerL 267
   | | | | | | | | | | | | | | | | | | | | | | | | | |
753 GCTCAAGCTGGAGCAGAGATCTACTTACCTTACCTTTCAGCTTTGATTCAT 802

```

```

267 eThrGlyProAlaHisTrpGlnValLeuLeuArgAlaArgAlaIleGlu 284
   | | | | | | | | | | | | | | | | | | | | | | | | | |
803 TACAGGCCAGCCCACTGGAGAGTGTGCTGGCGGCCCTGCTATGCATA 852
   | | | | | | | | | | | | | | | | | | | | | | | | | |
284 hGlnCysTyrrValValAlaAlaIleGlnCysGlyArgHisHisGlyLys 300
   | | | | | | | | | | | | | | | | | | | | | | | | | |
853 CCAGTGTATGTAGTGGCAGAGCAGCACTGTGTGAGAGCCACCATGAGA 902
   | | | | | | | | | | | | | | | | | | | | | | | | | |
301 ArgAlaSerTyrrGlyHisSerMetValValAspProTrpGlyThrVal 317
   | | | | | | | | | | | | | | | | | | | | | | | | | |
903 AGAGCAAGTATATGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGT 952
   | | | | | | | | | | | | | | | | | | | | | | | | | |
317 lAlaArgCysSerGlnGlyProGlyLeuCysLeuAlaArgIleAspLeu 334
   | | | | | | | | | | | | | | | | | | | | | | | | | |
953 GCGCCGCTCTCTGAGAGGCGCAGGCTCTGCTTGGCCCAATAGACTCA 1002
   | | | | | | | | | | | | | | | | | | | | | | | | | |
334 snTyrrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArg 350
   | | | | | | | | | | | | | | | | | | | | | | | | | |
1003 ACTATCTGGCAGAGTTGCCCGCAGACCTGCTGTGTCCAGACCGCAG 1052
   | | | | | | | | | | | | | | | | | | | | | | | | | |
351 ProAspLeuTyrrGlyAsnLeuGlyHisProLeuSer 362
   | | | | | | | | | | | | | | | | | | | | | | | | | |
1053 CCTGACCTTATGCAATCTGGGTGACCCACTGTCT 1088
   | | | | | | | | | | | | | | | | | | | | | | | | | |
seq_name: /SIDS1/gcgdata/hold-geneseq/gene-seq-nb1/NA2000.DAT:AAFI6257
seq_documentation block:
ID AAFI6257 standard; cDNA: 1382 BP.
XX
AC AAFI6257;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulvar; gastroenteric; nephrotoxic; antidiabetic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN M020005174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587513/55.
DR P-PSDB: AAB57054.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 1; Page 1124; 2338bp; English.
XX
CC AAFI5566 to AAFI6505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytosolic,
CC cardioactive, immunomodulatory, muscular, vulvar, gastroenteric,
CC nephrotoxic, antidiabetic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome

```


PR 05-AUG-1997: 97US-0055386.
 PR 05-AUG-1997: 97US-0055311.
 PR 18-AUG-1997: 97US-0055970.
 PR 18-AUG-1997: 97US-0055986.
 PR 19-AUG-1997: 97US-0056385.
 PR 19-AUG-1997: 97US-0056386.
 PR 19-AUG-1997: 97US-0056557.
 PR 19-AUG-1997: 97US-0056370.
 PR 19-AUG-1997: 97US-0056371.
 PR 19-AUG-1997: 97US-0056563.
 PR 19-AUG-1997: 97US-0056731.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Ebner R, Ferlie AM, Greene JW, Janat F, Ni J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
 XX
 DR WPI: 1999-167452/14.
 DR P-PSDB: AAY10877.
 XX
 PT New isolated human genes encoding secreted polypeptides - useful for
 PT diagnosis and treatment of pathological diseases

PS Claim 3; Page 265-266; 331pp: English.
 XX
 CC The specification describes secreted proteins and their corresponding
 CC polynucleotides which are useful for preventing, treating or ameliorating
 CC medical conditions, e.g. by protein or gene therapy. Pathological
 CC conditions can also be diagnosed by determining the amount of the
 CC secreted polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the products, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC diseases, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,
 CC prostate diseases, asthma, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
 CC thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.

CC
 XX
 SQ Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

alignment_scores:
 Quality: 1067.00 Length: 207
 Ratio: 5.205 Gaps: 0
 Percent Similarity: 99.034 Percent Identity: 99.034

alignment_block:
 US-09-357-675C-25 x AAX30398 ..

Align seg 1/1 to: AAX30398 from: 1 to: 1203

156 G1YLEuTPLeuSerLeuG1yG1yPheHISG1uArgG1yG1yAspTyrG1 172
 |||||||
 3 GGACTCTGGCTCTCCTGGTGGTGTTCATGACGCTGGCGCAAGACGCGGA 52
 |||||||
 172 uG1nThrG1nLys1eTyraSncYsHISValLeuLeuAAsnSerLySg1YA 189
 |||||||
 53 CGAGACTCAGAAATCTACAATTGTACAGCTGCTGCTGAACAG. AAAGGGG 101
 |||||||
 185 JAVa1ValAlaThrTyraG1yTyrHISLeuCYsAspValG1u1ePro 205
 |||||||
 102 CAGTGTGGCCACTTACAGAGACACATCTGTGACCTAGACATGAGATTCCA 151
 |||||||
 206 G1yG1nG1yPromeCYsG1uSerAsnSerThrMetProG1yProSer1e 222
 |||||||
 152 GGGCGAGGGGCT. ATGTGTAAAGCAACTTACCATGCTGGGCGCCAGTCT 200
 |||||||
 222 uG1uSerProValSerThrProAlaG1yLys1leG1yLeuAlaValCYST 239

|||||
 201 TGAGTCACCTGTACAGACACACAGCAGGCAAGATGCTACCTGTCTGCT 250
 |||||||
 239 YrAspMetArpPheProG1uLeuSerLeuAlaLeuAlaG1n1aG1yAla 255
 |||||||
 251 ATGACATGCGGTCCCTGAACCTCTCTGCGCAATGCGCTCAAGCTGAGGCA 300
 |||||||
 256 G1u1leLeuThrTyProSerAlaPheG1ySer1eThrG1yProAlaH1 272
 |||||||
 301 GAGATACCTACCTATTCCTTACAGCTTTTGGATCATTTACAGGCCAGCCCA 350
 |||||||
 272 sTrpG1uValLeuLeuAArgAlaArgAla1leG1uThrG1nCYsTYrValY 289
 |||||||
 351 CTGGGAGGTGTGTGCTGCGGGCCGCTGTATCGAACCCTGATGTGTAG 400
 |||||||
 289 a1Ala1a1a1aG1nCYsG1yArG1HISG1uLysArG1a1aSerTYrG1Y 305
 |||||||
 401 TGGCAGCAGCAGACAGTGTGGACGCCACATGACAGAAAGACCAAGTTATGGC 450
 |||||||
 306 H1sSerMetValValAspProTyrG1yThrValAla1a1aArG1sSerG1 322
 |||||||
 451 CACAGATGCTGTAGACCCCTGGGGAACAGTGTGGCCGCTGCTCTGA 500
 |||||||
 322 uG1yProG1yLeuCYsLeuAlaArg1leAspLeuAAsnTYrLeuArg1nL 339
 |||||||
 501 GGGGCGAGGCTCTGCTGCTGCGCAATAGACCTCAACTATCTGCGCAGACT 550
 |||||||
 339 euARgArG1s1euProValPheG1nHISArG1ArG1ProAspLeuTYrG1Y 355
 |||||||
 551 TCGCGCGACACCTGCTGTGTTCAGCACCGCAGCGCTGACCTTATGGC 600
 |||||||

356 AsnLeuG1yHISProLeuSer 362
 |||||||
 601 AATCTGGTCAACCACTGCTCT 621

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD25458

seq_documentation_block:

ID AAD25458 standard; cDNA: 1214 BP.

XX AAD25458;

XX 26-MAR-2002 (first entry)

XX Xenopus laevis Nrl1 cDNA.

XX Frog; Nrl1; cytosolic; neuroprotective; cellular pathway; therapy;

KW apoptosis; proliferative disorder; degenerative disease; ss.

XX Xenopus laevis.

XX Key Location/Qualifiers

FT CDS 180..1046

FT /*tag= a

FT /product= "Frog Nrl1 protein"

PD WO200187958-A2.

PD 22-NOV-2001.

PD 15-MAY-2001; 2001WO-US15664.

PD 16-MAY-2000; 2000US-204713P.

PD (UYJE-) UNIV JEFFERSON THOMAS.

PD Croce C, Brenner C, Pekarski Y;

PD WPI: 2002-082984/11.

PD P-PSDB: AAE15792.

PT Isolated cDNA encoding human, mouse, frog and yeast Nrl2 proteins,
 PT useful to find molecules that mimic or antagonize Nrl1 interaction for

PT the treatment of proliferative or degenerative diseases -
XX
PS Claim 21: Page 60; 61pp; English.

XX The invention relates to isolated nucleic acids comprising a fully
CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and
CC mouse Nlt2 proteins. Nlt and Fhit proteins are encoded as fusion
CC proteins in invertebrates and as separate polypeptides in vertebrates.
CC Nlt and Fhit interact physically and functions in same cellular pathways.
CC Molecules which bind Nlt2 and mimic or antagonise Fhit interaction are
CC used to treat diseases in which activity of Nlt2 protein is altered in
CC a mammal. Fhit mimics induce apoptosis and are particularly useful to
CC treat proliferative disorders, whilst Fhit antagonists promote cell
CC proliferation and are particularly useful to treat degenerative disease.
CC The present sequence is frog Nlt1 cDNA.

XX Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other:

alignment_scores:

Quality: 968.50 Length: 307
Ratio: 3.905 Gaps: 1
Percent Similarity: 80.782 Percent Identity: 57.980

alignment_block:

US-09-357-675c-25 x AAD25458 ..

Align seg 1/1 to: AAD25458 from: 1 to: 1214

51 LeucysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67
144 CTTGTGCGGATATAGGGGTAGAGCCGTGAGTCAATG..... 182
67 OArgProArgAlaMetAlaIleSerSerSerSerSerGlyLeuProLeuV 84
183GCTGGTCCACAAACCCCTGA 204
84 aAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelys 100
205 TTGCCGTGCGGATATAGGGGTAGAGCCGTGAGTCAATG..... 254
101 ThrCysAlaGlyLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAl 117
255 ACCTGTGCGGATATAGGGGTAGAGCCGTGAGTCAATG..... 304
117 aPheLeuProGlyAlaPheAspPheIleAlaArgAspProAlaGluThrL 134
305 GTTTCGCGGAAAGCCTTGACTATATCGGGGCGGACATTGAGAGAGAGC 354
134 eUnHisLeuSerGluProLeuGlyGlyLeuLeuGluGluGluGluGluGln 150
355 TGAAGTGTGCTGAGTCTTACATGGGACACCACTTCAGCGCTTACCCCAA 404
151 LeuAlaArgGlyCysGlyLeuThrPheLeuSerLeuGlyPheHisGluAr 167
405 CTGCGCAGGAGAGTGTGGCTTCCCTGGGGGATTTATAGAA 454
167 gGlyGlnAspTrpGluGlnThrGlnLysIleTyraAsnCysHisValLeuL 184
455 AGGACCAACCTGGACACGACCAACCACTTCCAAATTCACAGTGGTTG 504
184 euAsnSerLeuGlyAlaValAlaValThrTyraGlyThrHisLeuCys 200
505 TGAACACACAGGACATAGTATGCGTGTGACAGGCTCCTCGTTT 554
201 AspValGluIleProGlyGlnGlyPheMetCysGluSerAsnSerThrMe 217
555 GACGTAGACTTGCAGAAATGAGTGCATCAGAGAGAGCGATCCACCC 604
217 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 234
605 CCGCGGAGAGAGACTTATTCGCCCATCTTCACAGAGAAAGATTG 654

234 lYeuAlaValCysTyraSPMeLArgPheProGlnLeuSerLeuAlaLeu 250
655 GCGTGGGGGTGTGTACGACCTCCGCTCCCAATTCCTTGGGCTG 704
251 AlaGlnAlaGlyAlaGluIleuThrTyraProSerAlaPheGlySerI 267
705 GCCCAACAGAGACAGAACTTCACTTACCCCTTGCCTTACCCCTCAC 754
267 ePmGlyProAlaHisTrpGluValLeuArgAlaArgAlaIleGluTr 284
755 TACTGCTGCGCATTGGAGGTGTGCTGAGAGACCCGTCATYGA 804
284 hrcLncYsTyraValAlaAlaAlaGlnCysGlyArgHisGlyLys 300
805 CCGAGTCTATGATGTTGCGAGCGGACAGACAGACATGAGAG 854
301 ArgAlaSerTyrglyHisSerMetValAlaAspProTrpGlyThrVal 317
855 AGGACGCTCTATGTCACGCTATGCTAGACCCGTCGCGCTGCTCAT 904
317 lAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 334
905 TGCCCAATGCCAGAGAGAACAGAAATATGTATGCTGAGATTGACATTC 954
334 snTyLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArg 350
955 CTTACATGAGAGCTGTGAGCGGAGACATGCCGCTGTGAGAGCAGCAG 1004
351 ProAspLeuTyrglyAsnLeu 357
1005 ACTGATCTGTATGGAAATTC 1025

seq_name: /SIDSI/gcgcdata/hold-geneseq/geneseqn-emb1/MA2001B.DAT:ABLI2224

seq_documentation_block:

ID ABLI2224 standard; cDNA; 3548 BP.

AC ABLI2224;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.

KW Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, LI PMD, Myers EW;

DR WPI: 2001-655660/75.

DR P-PSDB; ABB68121.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

PS Claim 1: SEQ ID NO 31154; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB857737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;

alignment_scores:		
Quality:	715.50	Length: 364
Ratio:	2.944	Gaps: 6
Percent Similarity:	66.758	Percent Identity: 42.857

alignment_block:

US-09-357-675C-25 x ABL12224 . .

Align seg 1/1 to: ABL12224 from: 1 to: 3548

[illegible]

seq_name: /SIBS1/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL20222

seq_documentation_block:

ID ABL20222 standard; DNA; 5692 BP.

XX ABL20222:

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

OS

XX MO200171042-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX

XX Claim 1; SEQ ID NO 12139; 21pp + Sequence Listing; English.

[illegible][illegible]

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Claim 1: SEQ ID NO 31157; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;

alignment_scores:

Quality: 710.50 Length: 294
 Ratio: 3.259 Gaps: 3
 Percent Similarity: 74.150 Percent Identity: 48.639

alignment_block:

US-09-357-675c-25 x ABL12225

Align seg 1/1 to: ABL12225 from: 1 to: 1495

```

62 ValLeuCySLaGlnProArgProAlaMetAlaIleSerSerSe 78
||||:||||: ||| |||:||||:||||: |||
80 GTGATCGCATTCATCAGCAGTACGAGATGCTCCAAAAGAGAA 129
78 rCysGluLeuProLeuValAlaValCysGlnValThrSerTh 95
: |||: |||: |||: |||: |||: |||: |||
130 GGATCAAAAGTGCAGCAATTCGTGGCTGATGCCCTACACGCA 179
95 ySgInGlnAsnPhelYsThrCysAlaGluLeuValArgGlnAla 111
|||:||||: |||:||||: |||:||||: |||:||||: |||
180 AGCGGCGTAAATCTTACCAAGTATAGCTAGTATGATGAGCCAA 229
112 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspHei 128
: |||: |||: |||: |||: |||: |||: |||: |||: |||
230 CAGAACGCCCTGCATGCTCTTCTGCTGAGTGTGATTTGGGCGA 279
128 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyLys 145
: |||: |||: |||: |||: |||: |||: |||: |||: |||
280 GAGCCGACCCCAAACTATGAGCTCTCCGAGGGCTTGAGCGGAG 329
145 euGlnGluThrGlnLeuAlaArgGluCysGlyLeuThrLeuSer 161
: |||: |||: |||: |||: |||: |||: |||: |||: |||
330 TGCCGCAATACCGGAAATTCGCAAGTGCACAAATTTGGATTCC 379
162 gLyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 178
||||: |||: |||: |||: |||: |||: |||: |||: |||
380 GGTGGCGTCAGCAGCGAAGCAT.....CAAAAATCTT 414
178 rAsnCysHisValLeuLeuAsnSerYsGlyAlaValAlaThrTy 195
: |||: |||: |||: |||: |||: |||: |||: |||: |||
415 CAACGCTCATGTTTGTCTCAACGAGAAAGGGAAGTACGAGCAT 464
195 rGlyThrHisLeuCysAspValGluIleProGlyGlnGlyProm 211
||||: |||: |||: |||: |||: |||: |||: |||: |||
465 GAAAGCTGCACATGTTGATGATACGACTAA...GAGTTCGCC 511
212 GluSerAsnSerThrMetProGlyProSerLeuGluSerProVal 228
||||: |||: |||: |||: |||: |||: |||: |||: |||
512 GAATCAGATACAGTACGCGGATACCTGCTGAGCGCCAGTGAAG 561
228 rProAlaGlyLysIleGlyLeuAlaValCysTyTrpAspMetA 245
||||: |||: |||: |||: |||: |||: |||: |||: |||

```

```

562 TCAGATTGGCCAGATAGGCGCTCAGATTTCAGACCTCGTTTCTG 611
245 luleSerLeuAlaLeuAlaGlnAlaGluIleLeuThrTyPro 261
|||: |||: |||: |||: |||: |||: |||: |||: |||
612 AGCCGGCGGTGCTGCTAGAGAGCTGGTCCAAATTTGTTAACTAC 661
262 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeu 278
||||: |||: |||: |||: |||: |||: |||: |||: |||
662 TCCGATTCACATACGCAACCGGTAAAGCGCAGTGGGAATCTCTG 711
278 gAlaArgAlaIleGluThrGlnCysTyTrpValAlaAlaAlaGln 295
||||: |||: |||: |||: |||: |||: |||: |||: |||
712 GGCACAGCCATAGAGACTCAATGCTTTGTGCTGCGGCTCAGAT 761
295 lYargHisHisGluValArgAlaSerTyArgHisSerMetValAla 311
|||: |||: |||: |||: |||: |||: |||: |||: |||
762 GTTGCCACAAACACAGACGACAGATTTGGGCCACACATGATCCT 811
312 ProTrpGlyThrValValAlaAlaArgCysSerGluGlyPro... 327
||||: |||: |||: |||: |||: |||: |||: |||: |||
812 CCTGGGGAAGCTACTGCTGACTGCAGCAGCAGAGCTTGATATAG 861
327 sLeuAlaArgIleAsnLeuAsnTyTrpLeuArgGlnLeuArgHis 344
||||: |||: |||: |||: |||: |||: |||: |||: |||
862 CACGGCCGAGGTGACCTTCCGTCTCAATCTGTATCAGACCATWC 911
344 roValPheGlnHisArgArgProAspLeuTy 354
|||: |||: |||: |||: |||: |||: |||: |||: |||
912 CTTGCTTGAACATCTCTCGAAGACGACATCTAC 943

```

seq_name: /SIDSI/gcsgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAf84214

seq_documentation_block:

ID AAF84214 standard; cDNA; 1001 BP.

XX AAF84214;

DT 12-JUN-2001 (first entry)

DE Human hNt13-ase coding sequence.

KW Human; hNt13-ase; Cushings adrenal gland tumour;

KW nitrite hydrolytic enzyme; nitrilase; ss.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 99..929

FT /tag= a

FT /product= "Human hNt13-ase"

PN CN127998-A.

PD 27-DEC-2000.

PF 30-MAY-2000; 2000CN-0116221.

PR 30-MAY-2000; 2000CN-0116221.

XX (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.

XX Xu X, Qian B, Zhang X;

XX MPI: 2001-245678/26.

XX P-PSDB: AAB80984.

XX Human nitrite hydrolytic enzyme protein and its coding sequence -

XX Claim 1: Page 16-17; 20pp; Chinese.

XX The present invention relates to human hNt13-ase (nitrilase) protein,

XX which is expressed in Cushings adrenal gland tumour, and its coding

XX sequence (AAf84214 and AAB80984). The present invention also relates to a

CC preparation method of said protein and nucleic acid sequence, and a
 CC method of detecting human Nlt2-ase nucleic acid sequence and polypeptide
 CC in sample.

XX Sequence 1001 BP: 264 A; 235 C; 245 G; 257 T; 0 other;

alignment_scores:

Quality: 495.00 Length: 321
 Ratio: 2.403 Gaps: 8
 Percent Similarity: 64.174 Percent Identity: 34.891

alignment_block:

US-09-357-675C-25 x AAF84214

Align seg 1/1 to: AAF84214 from: 1 to: 1001

```

61 serValLeuGlyAlaGluProArgProArgAlaMetAlaIleSerSer 77
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|
11 GCCATTATGGCGGGGAGGAGCGAGCGCGCGCTGCTGATCTCCAGGCG 60
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
77 rSerCysGluLeu.ProLeuValAlaValCysGluValThrSer..... 91
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
61 CTCAGTCGCGCGCGAGGTGGTCTGCTGCTGACAGATCATGACCTCTTTC 110
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
92 .....ThrpAspGlyGlnGlnAsnPh 99
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
111 CGCTGGCCCTCATCCAGCTTCAGATTCTTCCTCAATGATTAACGT- 160
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
99 elYsthrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysL 116
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
161 CACTCGCGCTGTAGCTTCATCCGGAGGAGCAGCAAGCAGAGCAGCAAAA 210
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
116 euAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGlu 132
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
211 TAGTTCTTTGCGGAAATGCTT..... 233
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
133 ThrLeuHisLeuSerGluProLeuGlyLeuLeuGluGluTyr.. 148
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
234 .....AATCTCCATATGAGCGAATAATTTTCTCTGAATATGCG 271
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
149 .....ThrGlnLeuAlaArg 154
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
272 AGAGAAATTCCTGCGTGAATCCACAGAACCTTCTGCAATGACCAAGG 321
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
154 IucysGlyLeuThrPLeuSerLeuGlyPheHisGluArgGlyGlnAsp 170
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
322 AATGCACATATATCTCATTTGAGAGCTCTATCCCTGAA..... 359
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
171 TrpGluGlnThrGlnIleTyrAsnGlyHisValLeuLeuAsnSerLy 187
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
360 ...GAGGATGCTGGGAAATATATATACACCTGTGCTGTTGGGCGCTGA 406
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
187 sGlyAlaValAlaThrTyrArgGlyThrHisLeuGlyAspValGlu 204
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
407 TGGACCTTTACTACAAAGATATAGAAAGATCATCTGTTTACACTGATG 456
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
204 IeProGlyGlnGlyPLeuMetCysGluSerAsnSerThrMetProGlyPro 220
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
457 TTCTCTGGAATAATACATTTCAAGAACTAAACATGATGAGCGGGTGAT 506
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
221 SerLeuLeuSerProValSerThrProAlaGlyLeuGlyLeuAlaVal 237
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
507 AGTTTC...TCCACATTGTGATCTCTTACTGACAGAGTGGGTGGCGAT 553
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
237 IcyStyTyrAspMetArgPheProGluLeuSerLeuAlaValAlaGlnAla 254
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
554 CTGTAGACACTGGGTTTGCAGAGCTTGCACAAATCATACACACAGAGAG 603
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
254 IyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyPro 270
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
604 GCTCCAGAGCTGTTGTATATATCAGAGAGCTTTATCTGACACACAGAGCA 653

```

```

271 AlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluTrpGlnCysTyr 287
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
654 GCCCATTCGGAGTACTTACACGAGAACCGGCGCTGTGATATACAGTCTA 703
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
287 rValValAlaAlaIleGlnCysGlyArgHisHisGlnIleLysArgAlaSer 304
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
704 TGTGGCCACAGCTCTCTCTGCC.....CGGATGACAAAGCTCTCT 744
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
304 Yr.....GlyHisSerMetValValAspProTrpGlyThrValVal 317
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
745 ATGTTCCTGGGAGACACACCGGTGTGAACCTTGGGGGAGGTTCTTA 794
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
318 AlaArgCysSerGlnGlyProGlyLeuGlySerLeuAlaArgIleAspLeuAs 334
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
795 GCCAAGCTGGCAGAGAGAACATCTGTATTCAGACATAGACCTTAA 844
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
845 GAACTGGCTGAATAATACGCGCAAAATCCCGTTTATGACAGAAAGCAT 894
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
351 roAspPLeuTyr 354
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
895 CAGACCTCTTAT 905

```

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AA025456

seq_documentation_block:
 ID AA025456 standard; cDNA; 1359 BP.

AC AAD25456;

DT 26-MAR-2002 (first entry)

DE Human Nlt2 cDNA.

KW Human; Nlt2; cytosstatic; neuroprotective; cellular pathway; therapy;
 apoptosis; proliferative disorder; degenerative disease; ss.

OS Homo sapiens.

FT Key Location/Qualifiers
 CDS 22..852

FT /*tag= a
 /product= "Human Nlt2 protein"

PN W0200187958-A2.

PD 22-NOV-2001.

PF 15-MAY-2001; 2001MO-US15664.

PR 16-MAY-2000; 2000US-204713P.

PA (UYJE-) UNIV JEFFERSON THOMAS.

PI Croce C, Brenner C, Pekarski Y;

DR WPI; 2002-082984/11.

DR P-PSDB; AAE15790.

PT Isolated cDNA encoding human, mouse, frog and yeast Nlt2 proteins,
 useful to find molecules that mimic or antagonize Nlt2 interaction for
 the treatment of proliferative or degenerative diseases

PS Claim 3; Page 59; 61pp; English.

The invention relates to isolated nucleic acids comprising a fully
 defined cDNA nucleotide sequence encoding human, Xenopus laevis and
 mouse Nlt2 proteins. Nlt and Phlt proteins are encoded as fusion
 proteins in invertebrates and as separate polypeptides in vertebrates.
 Nlt and Phlt interact physically and functions in same cellular pathways.
 Molecules which bind Nlt2 and mimic or antagonise Nlt2 interaction are
 used to treat diseases in which activity of Nlt2 protein is altered in

CC a mammal. Flt1 mimics induce apoptosis and are particularly useful to
 CC treat proliferative disorders, whilst Flt1 antagonists promote cell
 CC proliferation and are particularly useful to treat degenerative disease.
 CC The present sequence is human Nit2 cDNA.

SO Sequence 1359 BP, 388 A; 284 C; 312 G; 374 T; 1 other;

alignment_scores:

Quality: 485.50 Length: 286
 Ratio: 2.542 Gaps: 6
 Percent Similarity: 66.783 Percent Identity: 36.713

alignment_block:

US-09-357-675c-25 x AAD25456 ..

Align seg 1/1 to: AAD25456 from: 1 to: 1359

```

84 ValAlaValAlcysGlnValThrSerThrProAspLysGlnGlnAsnPhely 100
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
37 TTGCGCCCTCATCCAGCTTCAGATTCTTCATCAATAACAGATACGTCAC 86
100 sThrCysAlaGlnValArgGlnAlaAlaArgLeuGlyAlaCysLeuA 117
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
87 TCGCGCTGTAGCTTCATCCGGAGCGACGACGCAAGCAAGCAATAAG 136
117 lApheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGluThr 133
   ::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
137 TTCTCTTGGCGGAATGCTTT..... 156
134 LeuHisLeuSerGlnProLeuGlyGlyLysLeuLeuGlnLutyr..... 148
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
157 .....AATTCTCATATGAGGCAATATTTCTCCATATATGAGA 197
149 .....  ThrglnLeuAlaArgGluC 155
198 GAAATTCCTGTGAATCCACACAGAGCTTTCTGAAAGTAGCAAGCAAT 247
155 ysGlyLeuThrPleuSerLeuGlyGlyPheHisGlnArgGlyGlnAspTrp 171
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
248 GCGCATATATCTCATTTGGAGGCTCATCCCTGAA..... 282
172 GluGlnThrGlnLysIleTyraNcysHisValLeuLeuAsnSerLysG1 188
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
283 GAGGATGCTGGGAAATTAATTAACACCTGCTGTGTTGGCCTGATGG 332
188 yAlaValAlaAlaThrTyraArgLysThrHisLeuCysAspValGluLeuP 205
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
333 AACTTACTAGCAAGATATAGAAAGATCCATTCGTTTGACATTGATGCTTC 382
205 roGlyGlnGlyPrometCysGluSerAsnSerThrmEtProGlyProser 221
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
383 CTGGAAATATTCATTCAAGAAATCTAATAACATTGAGTCCGGGTGATAGT 432
222 LeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCy 238
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
433 TTC...TCCACATTGATACTCTTACTGACAGAGTGGTCTGGCATCTG 479
238 sTyraMetArgPheProGlnLeuSerLeuAlaAlaGlnAlaGlyA 255
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
480 CTACGACATGCGGTTTGCAGAGCTTCACAAATCTACGCAACAGAGAGCT 529
255 lAluLeuLeuThrTyroSerAlaPheGlySerIleThrGlyProAla 271
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
530 GCGAGCTGTGGTATATCCAGAGCTTTTATCTGACCACTGGACCAAGCC 579
272 HisTrpGluValLeuLeuArgAlaAlaGlnAlaGlnLutyrGlyTyra 288
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
580 CATGTGAGATTACTTCAGCGAAGCGGCTGTGATATATCATGAGTATGTC 629
288 lValAlaAlaAlaGlnCysGlyArgHisGlnLysArgAlaSerTyrr. 304
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
630 GCGCAGAGCTCTCTGCTGCC.....GCGGATGACAAAGCCTCCTATG 670

```

```

305 .....GlyHisSerMetValAlaAspProTrpGlyThrValAla 318
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
671 TTGCTGTGGGACACAGCACCGGTGTGAACCTTGGGGGAGGTTCTTACGCC 720
319 ArgCysSerGlnGlyProGlyLeuCysLeuAlaAlaArgLLeuAsnTy 335
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
721 AAAGCTGGCACAGAGAACCAATCGTGTATTTCAGACATAGACCTGAAGAA 770
335 lLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProA 352
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
771 GCTGCTGAATATCGCCAGCAATCCCGTTTTCAGACAGACGATCAG 820
352 spleuTyrr 354
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
821 ACCTCTAT 828

```

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD25457

seq_documentation_block:

ID AAD25457 standard; cDNA; 1292 BP.

AC AAD25457;

DT 26-MAR-2002 (first entry)

DE Mouse Nit2 cDNA.

KW Mouse; Nit2; cytosolic; neuroprotective; cellular pathway; therapy;

KW apoptosis; proliferative disorder; degenerative disease; ss.

OS Mus sp.

FT Key Location/Qualifiers

FT CDS 46..876

FT /tag= a

FT /product= "Mouse Nit2 protein"

PN WO200187958-A2.

PD 22-NOV-2001.

PF 15-MAY-2001; 2001WO-US15664.

PR 16-MAY-2000; 2000US-204713P.

PA (UYE-) UNITV JEFFERSON THOMAS.

PI Croce C, Brenner C, Bekarski Y;

DR WPI: 2002-082984/11.

PS P-PSDB: AAE15791.

PT Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,

PT useful to find molecules that mimic or antagonize Flt1 interaction for

PT the treatment of proliferative or degenerative diseases

PS Claim 9; Page 59-60; 61pp; English.

XX The invention relates to isolated nucleic acids comprising a fully

CC defined cDNA nucleotide sequence encoding human, Xenopus laevis and

CC mouse Nit2 proteins. Nit and Flt1 proteins are encoded as fusion

CC proteins in invertebrates and as separate polypeptides in vertebrates.

CC Nit and Flt1 interact physically and functions in same cellular pathways.

CC Molecules which bind Nit2 and mimic or antagonise Flt1 interaction are

CC used to treat diseases in which activity of Nit2 protein is altered in

CC a mammal. Flt1 mimics induce apoptosis and are particularly useful to

CC treat proliferative disorders, whilst Flt1 antagonists promote cell

CC proliferation and are particularly useful to treat degenerative disease.

CC The present sequence is mouse Nit2 cDNA.

XX Sequence 1292 BP; 354 A; 292 C; 289 G; 357 T; 0 other;

alignment_scores: Quality: 481.00 Length: 302
 Ratio: 2.324 Gaps: 9
 Percent Similarity: 68.543 Percent Identity: 37.748

alignment_block:
 US-09-357-675c-25 x AAD25457 ..

Align seg 1/1 to: AAD25457 from: 1 to: 1292

```

59 GlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSe 75
   ||| |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7 CAGGATCCCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 rSerSerCysGlnLeuProLeuValAlaValCysGlnValThrSert 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49 ...TCTACTTCCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 91
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 hProAspIysGlnGlnAsnPhelYstHrcYsaIaGluLeuValArgGlu 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 TT.....AAATCAGATTAACCTTACCCTGGTGGCTGGGCTGGGCTGG 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109 AlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGlnAlaPheAsp 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GCGAGCAAGACAGAGTGGCCAAACATATGTTCTCTGCTGAGTGTTCAT 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 eIleAlaArgAspProAlaGluThrLeuHisLeu.....SerGluP 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....TCTCATATGACAAACCTTCTCTGCTGACTATGACAGAGA 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 rIleuGlyGlyLeuLeuGluGluGlyThrGlnLeuAlaArgGluCys 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 AGATTCTCTGGAGAGTCCACACAAAGCTTCTGAGATGACAAAGAGAG 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 GlyLeuThrLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrp 172
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 AGCATATATCTCATTTGAGGCTCCATCCCTGAA.....GA 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172 uGlnThrGlnLysIleTyraNcysHisValLeuLeuAsnSerLysGly 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 GGAATGCTGGGAAACTGTATATACCTGCTGTTGGGCTCGATGAGAA 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 IAlaValAlaThrTyraArgLysThrHisLeuCysAspValGluIlePro 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 GTTTACTGTGAAAGACAGAGATCATCTGTTGACATTTGATGATCTCT 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 GlyGlnGlyProMetCysGlnSerAsnSerThrMetProGlyProSerLe 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
409 GGGAAATTAACGTTTCAAGAAATCTAAACATTGAGCCCTGGTGTACTTT 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
222 uGlnSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCysT 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
459 C...TCCACATTGATACGCCCTTACGCAAAAGGGCCCTGGGATCTGCT 505
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 YrAspMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlyAla 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
506 ATGATATGCGCTTCGGGAGCTTCGACAAATCTATGACCAAAAGAGCTGC 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
256 GAlIleLeuThrTyProSerAlaPheGlySerIleThrGlyProAlaH 272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
556 CAGCTCTGTGTATCTCGAGCTTTCAATCTGACCCACAGAGACAGCCCA 605
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
272 sTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyraVal 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
606 CTTGGAGCTCTTCAGGAGCCCGGGCTGTGATATCAGGTGATGTGG 655
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289 aIaIaIaIaIaGlnCysGlyArgHisHisGluLysArgAlaSerTyra 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
656 CTCAGAGCTCTCTGCT.....CGGATGACAAAGGCTCTATGTG 696
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 .....GlyHisSerMetValAlaAspProTropolyThrValAlaIlaAr 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

697 GCCTGGGACACAGACACTGTTGTGATCTTGGGGCAGGCTTAACCAA 746
319 gCysSerGluGlyProGlyLeuCysLeuAlaIaArgIlaAspLeuAsnTyrl 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
747 AGCTGGCAGCAGAGGAAACAAATCTGTACTCAGACATCATGAGCTGAAGAC 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 eUaArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
797 TGGCTGAATTCGGCAGCAAAATCCCATTTTAAACAGAAACGACAGAC 846
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 LeuTyrr 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
847 CTCTAT 852
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

seq_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/AA2000.DAT.AAA81479
seq_documentation_block:
ID AA81479 standard; DNA; 69936 BP.
XX
AC AA81479;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; Menb; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
XX
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V,
PI Rappuoli R, Pizza M.
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 547-567; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AA81453 to AA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
CC AA81452 represent Neisseria meningitidis Menb polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariiae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.

```


CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.

XX
XX Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

alignment_scores:
Quality: 404.00 Length: 307
Ratio: 2.138 Gaps: 8
Percent Similarity: 61.564 Percent Identity: 35.179

alignment_block:

US-09-357-675C-25 x AAF21607 ..

Align seg 1/1 to: AAF21607 from: 1 to: 349980

51 LeucysProGluLeuArgIleProGluLeuSerValLeuCysAlaGlnPr 67
|||||
153960 TTGGTGTAAA.....CGTTTTCGCAAGTACCGTCATGTGTAGCGCTATCG 154003
67 OArg.....ProArgAlaMetAlaIleSerSerSerSerCysLeuPr 82
|||||
154004 GCGGAAATATGCCGACAGGTGCGCGCAAGAAAGAGAAAGAAATGCAAAA 154053
82 roleuValAlaValAlaCysGlnValThrSerThrProAspLysGlnGlnAsn 98
|||||
154054 TCAGAGTTGGCGCGCGTCAGATGTGTGCGCGCGTGTGCGCGCAACAC 154103
99 PheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCys 115
|||||
154104 GTCCGCCGCGCATGAACCGCTGTGCGCACGCGCGCGAGCGGTGCGGA 154153
115 sleuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 132
|||||
154154 TTGGGTCTCTCTCCGAAATATGTGTGTATGTCGTATGCGCGCAACGATACG 154203
132 IuThrLeuHisLeuSerGluProLeu...GlyGlyLysLeuLeuGlu 147
|||||
154204 ACAAACTCGGCGCTTGGCGAGCGCTTGGCGCGCGGACGCTTCAACAGCGCA 154253
148 TyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyLys 163
|||||
154254 TTGAGCGAAACGCGAAAGATGCGCGTGTGCTG...TTCGCGCGGAGC 154300
164PheHisGluArgGlyGlnAspTrpGluInThrGln 176
|||||
154301 TGTCCGCTGCAAAAGTGTGCGAGCGGT.....A 154329
176 ysIleTyTrpAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 192
|||||
154330 AAGTGATGAAATACGCTGTGTGTACGACGCGACGCGCGTGAAGACGCGG 154379
193 ThrTyArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 209
|||||
154380 CTGTACCCACAAATATGCACCTC.....TTCGGTTTTCGCGTTGTGGCGA 154423
209 oMetCysGluSerAsnSerThrMetProGlyProSerLeuSerLeuSerProv 226
|||||
154424 ACGCTATGCGCAAGCGATACCATCCGCGCGGCGGATGTGCCGCACT 154473
226 alSerThrProAlaGlyLysIleGlyLeuAlaValAlaCysTyAspMetArg 242
|||||

154474 TGTGCGCAGAAAGCGGTGCGGTCGCGGCGGCGCATTTGTTACGATGTCGCC 154523
243 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 259
|||||
154524 TTTCGCCGAA.....TTTTTCGACGCGAGTTCGCGTTGTGACGTAATTAT 154567
259 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 276
|||||
154568 GCTGCCCGCTCGTTTACGACACGACGCGGCGCAAGCGCATTTGGAGCTGC 154617
276 euleuArgAlaArgAlaIleGluThrGlnCysTyValAlaAlaAla 292
|||||
154618 TGTGCGCGCGCGTGTGCGCAAAACCAATGTGTACGTGCGCGCGCGCA 154667
293 GlnCysGlyArgHisHisGlnLysArgAlaSerTyTrpGlyHisSerMetVal 309
|||||
154668 CAGGCGCGTTTGCACGAAACGAGCGCGCGCATGTCGCGACACGATAT 154717
309 lValAspProTrpGlyThrValAlaAlaArgCysSerGluGlyProGlyL 326
|||||
154718 TGTGATCCGTGTGCGCGCACGTGTGTGACGTATTCGCCGAGGCGCAAGCG 154767
326 euCysLeuAlaArgIleAspLeuAsnTyTrpLeuArgGlnLeuArgArgHis 342
|||||
154768 TTGTACGCGACAGATCGATGCGCAACCGCTGAAACAGCGTCGCAACCGC 154817
343 LeuProValPheGlnHisArg 349
|||||
154818 CTGCCCGCTTGAAATACCGG 154838
seq_name: /SIDSL/gcgdata/hold-sequences/geneseqn-emb1/NA2000.DAT:AAA81490
seq_documentation_block:
ID AAA81490 standard; DNA; 1437668 BP.
XX
XX AAA81490:
AC
XX
XX 04-DEC-2000 (first entry)
DT
XX
XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
DE
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
XX Neisseria meningitidis.
OS
XX
XX MO200022430-A2.
PN
XX
XX 20-APR-2000.
PD
XX
XX 08-OCT-1999; 99MO-US23573.
PF
XX
XX 09-OCT-1998; 98US-0103794.
PR
XX
XX 30-APR-1999; 99US-0132068.
PA
XX
XX (CHIR) CHIRON CORP.
PI
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignan V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;
PI Rappuoli R, Pizza M;
XX
XX WPI; 2000-318079/27.
DR
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7; Page 866-1272; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA

sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding primers; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes; and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

alignment_scores:
Quality: 404.00 Length: 307
Ratio: 2.138 Gaps: 8
Percent Similarity: 61.564 Percent Identity: 35.179

alignment_block:
US-09-357-675C-25 x AAA81490 ..

Align seg 1/1 to: AAA81490 from: 1 to: 1437668

```
51 LeucylProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 67
|||||
453960 TTGTTGTA...CGTTTCGCAAGATACGTCATGCTACCTATCG 454003
|||||
67 GATG...ProArgAlaMetAlaIleSerSerSerSerCysGluLeuPr 82
|||||
454004 GCGGAATATGCGCCAGAGTCGCGCAAGAAAGAGAAATGACAAA 454053
|||||
82 roLeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsn 98
|||||
454054 TCAGAGTTGCGCGCGTCAGATGTCGCGCGTCGCGCGGAACAC 454103
|||||
99 PheLysThrCysAlaGlnLeuValArgGluAlaAlaArgLeuGlnLac 115
|||||
454104 GTGCGCGCATGAAACGCGTGTGCGCGCGCGCGAGCGAGGTGCGGA 454153
|||||
115 IleuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 132
|||||
454154 TTGGGTGCTGCTCCCGAATATTGGGTGCTGATGGCGCAAGATACCG 454203
|||||
132 IuThrLeuHisLeuSerGluProLeu...GlyLysLysLeuLeuGln 147
|||||
454204 ACAAACTCGGCTTGGCGAGCTTTGGCGCGCGAGCTTTCAGACGCA 454253
|||||
148 TyrThrGlnLeuAlaArgLysGlyLeuThrLeuSerLeuGlnGly.. 163
|||||
454254 TTGAGCGAAACGGCGAAAGATGCGGTGTCCTG...TTTCGCGGAC 454300
|||||
164 .....PheIleGluArgGlyGlnAspTrpGluGlnThrGlnL 176
|||||
454301 TTGCGCGCTGCAAAAGCTCGAGCGGCT.....A 454329
|||||
176 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 192
|||||
454330 AAGTGATGATACGCTGTTGTCACGACGAGCGCGTAAGACGGCG 454379
|||||
193 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 209
|||||
```

```
454380 CTGTACCAAAAATGACACCTC.....TTGGTTTTCGCGTTGGCGCA 454423
|||||
209 omelCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 226
|||||
454424 ACCGTATGCCGAAGCCGATACATCCGCGCGCGGAGATGTCGCACT 454473
|||||
226 alSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 242
|||||
454474 TTGCGGAGAAAGCGTCCGTCGCGCGCGGCAATTGTCATGATGTCGC 454523
|||||
243 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeu 259
|||||
454524 TTTCGCGAA.....TTTTCGACGCGCACTGCTTTCGATTTGAT 454567
|||||
259 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 276
|||||
454568 GCTGCCGCTGCTTTCGACAGCGGCGCAAGCGCATTTGGAGCTGC 454617
|||||
276 euleuArgAlaArgAlaIleGluThrGlnCysTyrValAlaAlaAla 292
|||||
454618 TGCTGCGCGCGCGTCCGTCGCAAAACCAATGTTACGTCGCGCGCGCA 454667
|||||
```

293 GlnCysGlyArgHisHisGlyLysArgAlaSerTyrCylHisSerMetVa 309

454668 CAGGCGGCTTTCGACGCAAGAGCGCGCACTTCGACAGACAGAT 454717

309 lValAspProTrpGlyThrValAlaAlaArgCysSerGlnGlyProGlyL 326

454718 TGTCGATCCCTGGGCGGCGACGTCGTCGACGTAATTCGCGAGCGCAAGCGC 454767

326 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgHis 342

454768 TTGTTACGCGACAGCATGATGCCACCGCTGAACAGCGTCGCAACCGC 454817

343 LeuProValPheGlnHisArg 349

454818 CTGCGCGCTTGAATACCGG 454838

seq_name: /SIDSI/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC77176

seq_documentation_block:

ID AAC77176 standard; cDNA; 1725 BP.

```
AC AAC77176;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF2731 polynucleotide sequence SEQ ID NO:5461.
DE
XX
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
XX vulnerability; antiproliferative; antiparkinsonian; neuroprotective;
XX anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiac;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antineoplastic;
XX antiviral; antibacterial; antifungal; antihypertensive; antihypertoid;
XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
```



```

PR   31-MAR-1999;       99US-0127607.  
PR   02-APR-1999;     99US-0127636.  
PR   05-APR-1999;     99US-0127728.  
XX   30-MAR-2000;    2000US-0540763.  
  
PA   (CURA-) CURAGEN CORP.  
PI  
XX   Shinkets RA, Leach M;  
DR  
XX   WPI: 2000-602362/57.  
DR   P-PSDB; AAB42967.  
PT  
XX   Novel nucleic acids and peptides derived from open reading frame X,  
PT   useful for treating e.g. cancers, proliferative disorders,  
XX   neurodegenerative disorders and cardiovascular disease -  
PS  
XX   Claim 5; Page 4641-4643; 5507bp; English.

AAC77446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antiparasitic; antiParkinsonian; nootropic; neuoprotective; osteopathic; anticoagulant; antifibrillic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic; antididiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antineumatic; antihypoid; and antinaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypocholelrdism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease, to enhance coagulation; to inhibit thrombolysis; and as a contraceptive.



Sequence 1725 BP; 466 A; 368 C; 372 G; 517 T; 2 other;



Alignment scores:



|                     | Quality:     | Length:                  |
|---------------------|--------------|--------------------------|
|                     | Ratio: 1.858 | Gaps: 14                 |
| Percent Similarity: | 43.867       | Percent Identity: 24.324 |



alignment block:  
US'09-357-675C-25 x AAC77176 ..



Align seg 1/1 to: AAC77176 from: 1 to: 1725



```

58 ProGlnIeuSeRVallencysAlacInProArGrPrOAcylametaAlalII 74
 |||||::: ::::::|||
11 CGCGAGtGTCTTcTGCC.....AAgGTATG.... 40
74 eSerSerSercyGIuleuProLeuValalaYalcysglInvalThrs 91
 :||| |::|| |::| |::: |:|:|||||
41ACTCTTCGCCGGTGCCTCATCAAGTT...CAGAATTCTT 80
91 erThrProAsplyscIngInasnpheLyfThCySaIGluLeuValArg 107
 || |:::||| |:::| |:::| |::| |::|
81 CCATC.....AAATCACATAACGTCACTCGGCTGTGAAGCTTACC GG 124
108 GUAlaalaaargLcuGlYalacyScyuLaPhelacuProGLialaphas 124
 ||||| |::|| |::| |::| |::| |::| |::|
125 GAAGCAGCAACCAAGAGCAAAAATAgTTTTGTCGGAGATGCTTT.. 172
124 pHeilleaIarGaSProlaIgLUthrLeuhIsleuSerGIuprOLEug 141
173ATTCGTCATNG 185
```


```

141	lyglylysluLeuLenglugluytr.....	146
146	gagcccaaaatatttttccctaaataatgcagagaaaattcctggtgaattcacaca	235
149ThglnleuAlaArglucyluGlyleuTrpLeuSerLeuGl	162
236	cagaaaccttttcgaagctagcacagaaagaaatgcacatatatctcatttggagc	285
162	lylypHe-HisgluArglyln...AsptrpGluglnThrGlnIysile	177
286	taacttctcaaccacaaagagctctatccctgaagaaagatgcgggaaattta	335
178	tyrAsnCysHsIvalLeuLeuAsnSerIysglYalaIvalAlaIaThrTy	194
336	tataacaccctgcctgctgtttggcgccatgagaaactttactagcAAAGTa	385
194	trArglyThrHisIeuCysAspValGlnIleproglylnGlyPrometc	211
386	tagaaagatccatcgttcttgacatgagcttctggaaaaatttaccatttc	435
211	ysgluSerAsnSerIhrmetProglyProSerLeuGlu.....	223
436	agaaTCTTAAACATCTGAGTCCGGGTGATAGTTCCTCCACATTGATACT	485
223	223
486	cgtatgtaccagatAAGTTGGCTCTTAgcAATCTcAGTAgAAgACAAT	535
224SerProValSer.....	227
536	cagTATTTATTCTTTTTTGTCTCTCCGATTTCTTCACATAACTTAA	585
227	227
586	ctgAAAGACCATAACTGAGAAAGGACAGAAATCATACAGACTCTGGAAG	635
227	227
636	ttcgggcttatttgagaaCTAAGATATTGACAGATTTTGCCTTTGATT	685
227	227
686	tgatttgacTTCtGTTAACGgCTTCAGAGTAACTATTAGGCTACAG	735
227	227
736	ttgagTACCTCCACTGATGATATAAGATTCATTAAGATGATTCTC	785
228ThrProAla.GlyLys.....	232
786	atctttactccgcctgatGTAATGATGTCttTATGATGAAGTCCAAGT	835
232	232
836	aggaATGAGCTTGTAAATTATCTGCTCTCAGGTCtGTGTTAATTAT	885
232	232
886	ccctgcTCAgTtTTTGTATCTATATGTCATGAGAGATTCCCTGCAC	935
232	232
936	accATGCTGTAGGAGATTAACTTTTCATTTCGATTTCTGTTGGAA	985
233IleGlyLeuAlaValCysTyrAspMetArgpHePro	244
986	cagCTTACTGCAAGTGGCTCGGCATCTGCTAGACATGCGGTTTCA	1035
245	gluLeuSerLeuAlaLeuAlaGlnIaGlyAlaGluIleLeuThrTyrPr	261
1036	gagCTTGCACAAATCTAGCACAGAGAGGCTCCAGCTGTTGTTATATCC	1085
261	oSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuA	278


```

312 ProTrrpGlyThrValAlaAlaArgCysSerGluGlyProGlyLeuCysLeu 328
|||||::: ||| ::||| ||| :::
746 CCATGGGCCCAAGCTGCAGACAGTGCCTGAAGCGCAGGAATCGTGT 795
328 uAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuProy 345
::||| |||::: ::: |||::: |||::: |||
796 GGCCGATATGATTTCTCCGAGGTGGAGACAGTGCCTCAGCAGATTCCTCCG 845
345 aPheGlnHisArgArgProAspLeuTyr 354
||||| ::||| ||| |||
846 TCTTTGGGCAAGACGCTCTAGATCTGTAC 874

```

THIS PAGE BLANK (USPTO)

Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,
URL: <http://genome-gsc.riken.go.jp/>, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (<http://genome-gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGGAGAGAGCGCGCCACACCTCGACTTTTCTTTTNN 3'], cDNA was
prepared by using triethanolamine-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0. Second strand cDNA
was prepared with the primer adapter of sequence [5'
GAGGAGAGAGAGATCCAGATCAATTAATTAATTAACCCCGCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

FEATURES
source Location/Qualifiers

1..1374
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGI:1891931"
/db_xref="taxon:10090"
/clone="061007B07"
/sex="male"
/tissue_type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..1374
/gene="Ntli"
1..1374
/note="data source:MGI, source key:MGI:1350916,
evidence:ISS
ntrilase 1"

BASE COUNT 330 a 359 c 342 g 343 t
ORIGIN

alignment_scores:
Quality: 1529.00 Length: 364
Ratio: 4.662 Gaps: 4
Percent Similarity: 90.110 Percent Identity: 81.044

alignment_block:
US-09-357-675C-25 x AK002269 ..

Align seg 1/1 to: AK002269 from: 1 to: 1374

```
1  ProLeuAlaAlaAla**LeuAlaProAspArgProAspArgThrLe 17
|||||||  |||  |||||||  |||  |||
10  CCGCTGCTGCTGCGCTGCGCTGCGAGCTGTGAGATCGACCC 59
17  uArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 34
|||||||  ::::  |||||||  |||  :::
60  GTGAATGTTGGCTGTGCTTCATAGAGACTACTCCCAACCTGC 109
34  rArgPro.ArgLeuGlyPheIleThrArgProProHisArgPheLeuSer 50
|||||||  |||||||  |||||||  |||  :::
110  CCGCTGCGGCTGCTCATCACAGGCTCTCCACCA.....CT 150
50  uLeuCyProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnP 67
|||||||  |||  ::::  |||||||  |||  :::
151  CCGTGTACCGGATACGATTACTTCGACCCCGACTTTGTACTGAC 200
67  rArgProArgAlaMetAlaIleSerSerSerCysGlnLeuProLeu 83
|||||||  ::::  |||||||  |||  :::
201  CCAAGGCCCAACCATGTCC...TCATCACTCTCGGAGGCTCCCTG 247
84  ValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhely 100
|||||||  ::::  |||||||  |||  :::
248  GTGGCTGTGTGCGAGTACATCAACCAACCAAGCAAGAACTTAA 297
```

```
100  sThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeu 117
|||||||  ::::  |||||||  |||  :::
298  AACATGTCTCAGTGGTTGACAGAGGCTGCCAGACTGGTGTCTGCTCG 347
117  LaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThr 133
|||||||  ::::  |||||||  |||  :::
348  CTTTCTGCTGCGAGCACTTTCATCTTATTCACCAAACTTCGCCGAGACA 397
134  LeuHisLeuSerGluProLeuGlyGlyLeuLeuGluGluThrThrGlu 150
|||||||  |||||||  |||||||  |||  :::
398  TTACTCTGTCGACCAACCACTGAATGGGATCTTTGGGCCAATATACCA 447
150  nLeuAlaArgGluCys.GlyLeuThrLeuSerLeuGlyGlyPheHisGlu 166
|||||||  |||  ::::  |||||||  |||  :::
448  GCTTCCAGGGAATGTGGAATCGGCTGCTGCTGGCGGTTCCACAGAG 497
167  ArgGlyGlnAspThrGluGlnThrGlnLysIleThrAsnCysHisValLe 183
|||||||  ::::  |||||||  |||  :::
498  CGTGCCCAAGACTGCGAGCAGAAATCAAAATCTACATTCATGTGCT 547
183  uLeuAsnSerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeu 200
|||||||  |||||||  |||||||  |||  :::
548  TTGAACAGCAGAGGATCAGTAGTGGCCAGTTACAGAGACACATCTGT 597
200  ysAspValGluIleProGlyGlnGlyProMetCysGlnSerAsnSerThr 216
|||||||  |||||||  |||||||  |||  :::
598  CGGATGTAGAGATCCACAGTCAGGCGCGATGACAGAAACACATATACC 647
217  MetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 233
|||||||  ::::  |||||||  |||  :::
648  AAGCTCGAGGACACTTGGAGCCACTGTCAAGACACCGCTGCGCAAGT 697
233  eGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAla 250
|||||||  ::::  |||||||  |||  :::
698  TGGCTAGCAATCTGTATGACATGCGGTTCCGAACTTCTTGAAT 747
250  euAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySer 266
|||||||  |||||||  |||||||  |||  :::
748  TGGCTCAAGCTGGGCGAGAAATACTTACTTATCTTACGCTTATGATCT 797
267  IleThrGlyProAlaHisThrGluValLeuLeuArgAlaArgAlaIleG 283
|||||||  ::::  |||||||  |||  :::
798  GTTACAGCTCCAGCCCACTGGAGTGTCTGTGGGCGCGCATTTGA 847
283  uThrGlnCysTyrValAlaValAlaAlaGlnCysGlyArgHisHisGlu 300
|||||||  |||||||  |||||||  |||  :::
848  ATCTCAGTGTCTATGTATATACGACGACGCACTGTGGACGCCACCATG 897
300  ysArgAlaSerTyrGlyHisSerMetValValAspProThrGlyThrVal 316
|||||||  |||||||  |||||||  |||  :::
898  CAAAGCAAGATTATGGCCATAGCATGTGGTTGACCCGTGGGGCAGACTG 947
317  ValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLe 333
|||||||  ::::  |||||||  |||  :::
948  GTGGCCCGCTGCTCCGAGGAGCAGGCGCTGCTTCTCGAATTTGATCT 997
333  uAsnTyrLeuAlaArgGlnLeuArgArgHisLeuProValIlePheGlnHis 350
|||||||  |||||||  |||||||  |||  :::
998  CCACTTCTACAAACAGATGCGCCAAACACTGCTGTCTTCAGACCCGCA 1047
350  rGProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362
|||||||  |||||||  |||||||  |||  :::
1048  GACCTGACCTCTATNGGACAGTGTGGGTATATCAGTCTCT 1085
```

seq_name: gb_hlc:AK004988
seq_documentation_block: 2993 bp mRNA linear HTC 19-JAN-2002
LOCUS AK004988
DEFINITION Mus musculus adult male liver cDNA, Riken full-length
library, clone:1300013f05:ntrilase 1, full insert sequence.
ACCESSION AK004988
VERSION AK004988.1 GI:12836590


```

113 yalaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAsp 130
129 TGCTTGCTGCTGCTTTCCTGCTGAGCATTTGACTTATTGCAGAAACC 340
130 rolaAGlUthrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGlu 146
131 CTTGGCGAGCATTAATCTCTCCGACCACTGAAATGGGATCTTTTGGCC 390
147 GlUthrGlnLeuAlaArgGluCysGlyLeuThrPheLeuSerLeuGly 163
148 CATTATAGCCAGCTTCCAGGAGATGTGGAATCTGCTGCTTGGCCGG 440
163 yPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsn 180
164 TTTCCAGCAGCGTGCCAGACTGGAGCAGAAATCAGAAATATCAAT 490
180 yHisValLeuLeuAsnSerLysGlyAlaValAlaIleThrTyrArgLys 196
181 GTCATGTGCTTTTCACAGCAGGATCAGTGTGCTGCTGCTGCTGCTG 540
197 ThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGlu 213
198 ACACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 590
213 rAsnSerThrMetProGlyProSerLeuGluSerProValSerThrPro 230
214 CAATATACCAAGCCTGAGGACTTGTGAGCCACTGTCAAGACACCGG 640
230 laGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeu 246
231 CTGGCAAGGTGTGCTGCAATCTGTATGACATGCGTCCCTGCAACT 690
247 SerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSer 263
248 TCTTTGAAATTTGCTCAAGCTGGGCGAGAAATCTTACTTATCTTAC 740
263 apheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAla 280
264 CTTTGATCTGTGTACAGTCCAGCCACTGGAGGCTGCTGCTGCTGCT 790
280 rglAlaIleGluThrGlnCysTyrValAlaAlaAlaGlnCysGlyArg 296
281 GCGCATATGATCTCAGTGTATGATATGACAGCAGCAGTGTGAGCG 840
297 HisHisGluLysArgAlaSerTyrGlyHisSerMetValAlaAspPro 313
298 CACCATTAACAAGACAGCAAGTATGAGCCATGATGCTGCTGCTGCTG 890
313 rglYthrValAlaAlaArgCysSerGluGlyProGlyLeuCysLeuAla 330
314 GGGCAGCAGTGTGGCCGCTGCTCCAGGAGCAGGCTGCTGCTGCTG 940
330 rglLeaAspLeuAsnTyrLeuArgGlnLeuArgAlaHisLeuProValPhe 346
341 GAATGTATCTCCACTTTCTACAAAGATGGCGCAACACCTGCTGCTT 990
347 GlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362
348 CACGACCGCAGACCTGACTGTATGAGCAGTGTGGGTATCATCCTCT 1038
391 CACGACCGCAGACCTGACTGTATGAGCAGTGTGGGTATCATCCTCT 1038

seq_name: gb_est1:AL520768

seq_documentation_block:
LOCUS AL520768 890 bp mRNA linear EST 13-FEB-2001
DEFINITION AL520768 LTI_NFL004_NBC2 Homo sapiens cDNA clone CSDB0021D05 5
prime, mRNA sequence.
ACCESSION AL520768
VERSION AL520768.1 GI:12784261
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 890)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.

FEATURES
source
1..890
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDB0021D05"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact: Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax: (1) 301 610
8371 Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com"

BASE COUNT 191 a 251 c 241 g 206 t 1 others
ORIGIN

alignment_scores:
Quality: 1524.00 Length: 306
Ratio: 5.201 Gaps: 1
Percent Similarity: 95.752 Percent Identity: 95.752

alignment_block:
US-09-357-675C-25 x AL520768 ..
Align seg 1/1 to: AL520768 from: 1 to: 890

7 LeuAlaProAspArgProProAspArgThrLeuArgMetValLeuAlaI 23
|||||
4 CTGGCTCCAGACCGCCCTCCGATCGACCCCTGGGAATGGTTGGCTAT 53
23 eSerSerCysArgThrTyrSerLeuSerArgArgProArgLeuGlyPhe 40
|||||
54 ATCTTCATGCTG.....GGCTCA 72
40 leThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLeuArg 56
|||||
73 TCACACAGCCTCCTCAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCT 122
57 IleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAl 73
|||||
123 ATACCTCAACTCTCAGTACTTGTGTGCTCAGCCCGCCAGCGCAGCGCAT 172
73 alleSerSerSerSerCysGluLeuProLeuValAlaValCysGlnVal 90
|||||
173 TATCTCTTCTCTCTCTGCGAAGTGGCCCTGGTGGTGTGTGGCGGTA 222
90 hrSerThrProAspLysGlnGlnAsnPheLysThrCysAlaGluLeuVal 106
|||||
223 CATTCAGCGCCAGACACAGCAAGCAACTTTAAACATGTGCTGAGCTGTT 272
107 ArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaP 123
|||||
273 CGAGAGGCTGCCAGACTGTGGTGGCTGCTGCTGCTGCTGCTGCTGCT 322
123 eaSPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProL 140
|||||
323 TGACTTCATTGACAGGAGCCCTGACAGAGCATTACACTGTGTGAACAC 372

```



```

140 euglyglyleuLeuGlulGlyThrGlnLeuAlaArgGlyCysGly 156
|||||
373 TGGGTGGGAACCTTTGGAGAAATACACCCAGCTTGGCCAGGAATGTGGA 422
157 LeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluG 173
423 CTCGTGGCTGCTGGGTGGTTCATGACAGCGTGGCCAGACTGGAGACA 472
173 nHrGlnLysIleTyraAsnCysHisValLeuLeuAsnSerLysGlyAla 190
473 GACTCAGAAAATCTACAATTGTCAAGCTGCTGTAACAGCAAGAGGGCAG 522
190 aValAlaThrTyraGlySerThrHisLeuCysAspValGluIleProGly 206
523 TAGTGGCCACTTACAGAGACACATCTGTGTGACGTAGAGATTCCAGGG 572
207 GlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerLeuG 223
573 CAGGGGCGCTATGTGTGAACCACTTACCATGGCTGGCCCACTCTTGA 622
223 uSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCysTyra 240
623 GTCACTGTGACACACACAGCAGCAAGATTGCTAGCTGTCTATG 672
240 spMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGlu 256
673 ACATGGCGTTCCCTGACACTCTCTGTGCATGGCTCAAGCTGGAGCAGAG 722
257 IleLeuThrTyProSerAlaPheGlySerIleThrGlyProAlaHisTr 273
723 ATACTTACCTATCCTTCAGCTTTGGATTCATTACAGGCCAGCCCACTG 772
273 pGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyraValAla 290
773 GGAGGTGTCTGTGGGCGCGTGTATCGAAACCCAGTGTCTATGTAGTGG 822
290 IaIaIaGlnCysGlyArgHisHisGlnLysArgAlaSerTyrglyHis 306
823 CAGCAGACAGTGTGGACGCCACCATGAGAAAGAGACAAAGTATTGGCCAC 872
307 SerMetValValAspPro 312
873 AGCATGGTGGKAGACCCC 890

seq_name: gb_est1:AL529152

seq_documentation_block:
LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529152 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSDD004YM06 5
prime, mRNA sequence.
ACCESSION AL529152
VERSION AL529152.1 GI:12792645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 844)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Source 1..844
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CSDD004YM06"
/clone_id="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"

```

```

/!ab host="DH10B"
/!note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by life technologies. Contact: Feng Liang life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 182 a 238 c 235 g 189 t
ORIGIN

alignment_scores:
Quality: 1413.00 Length: 263
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-357-675C-25 x AL529152 ..
Align seg 1/1 to: AL529152 from: 1 to: 844

100 LysThrCysAlaGluLeuValAlaGluAlaAlaArgLeuGlyAlaCysLe 116
|||||
1 AAAACATGTGCTGAGCTGGTTGAGAGGCTGCACAGCTGGGTGCTGCTGCT 50
116 uAlaPheLeuProGlnAlaPheAspPheIleAlaArgAspProAlaGly 133
51 GCGTTTCTGCTGAGCATTTGACTTATTCACAGGAGACCCTGCAGAGA 100
133 hIleuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGlyTrh 149
101 CGCTACACCTGCTGAAACCACTGGGTGGAAACTTTGGAAGAAATACAC 150
150 GlnLeuAlaArgGlyCysGlyLeuTrpLeuSerLeuGlyGlyPheHisG 166
151 CAGCTGGCAGGGAATGTGACTGTGCTGTGCTGGGTGGTTCATGAGA 200
166 uArgGlyGlnAspTrpGluGlnThrGlnLysIleTyraAsnCysHisVal 183
201 GCGTGGCCAGACTGGAGACAGACTCAGAAAATCTACAAATTGTCTACG 250
183 euleuAsnSerLysGlyAlaValAlaIaIaThrTyraGlySerHisLeu 199
251 TGTGAAACAGCAAGAGGCGAGTAGTGGCCACTTACAGAAAGACACATG 300
200 CysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerTh 216
301 TGTGACGTAGAGATTCCAGGGCAGGGGCTGTATGTGAAACCACTTAC 350
216 rMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 233
351 CATGCTTGAGGCTTTGAGTCACTGTCAAGCACACAGCAGGCAAGA 400
233 leGlyLeuAlaValCysTyraAspMetArgPheProGluLeuSerLeuAla 249
401 TTGGTCTAGTGTCTGCTATGACATGCGGTTCCCTGCAACTCTCTGCA 450
250 LeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAlaPheGlySe 266
451 TTGGTCAAGCTGGAGCAGAGATACTTTCATCTTACGCTTTTGGATC 500
266 rIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaIaGly 283
501 CATTACAGGCCAGCCCACTGGAGAGTTTGTCTGGGCCCCGTGCTATCG 550
283 luthrGlnCysTyraValAlaAlaAlaGlnCysGlyArgHisHisGlu 299
551 AAACCACTGCTATGTAGTGGCAGCAGCAGTGTGAGCCACCATGAG 600

```

```

300 LysArgAlaSerTyrGlyHisSerMetValAlaAspProTrrpGlyThrVa 316
|||||
601 AAGAGCAAGCAAGTTATGGCCACAGATGTGTAGACCCCTGGGGAAACAGT 650
|||||
316 ValAlaIaArgCysSerGluGlyProGlyLeuCysLeuAlaIaArgIleAspL 333
|||||
651 GGTGGCCCGGTGCTGTGAGGGGGCAGGCCCTGTGCTTGGCCGAAATAGACC 700
|||||
333 euaSnTyrLeuArgGlnLeuArgAlaArgHisLeuProValPheGlnHisArg 349
|||||
701 TCACATATCTGGCCACAGTTGCGCCGACACCTGCTGTGTTCCAGACCCG 750
|||||
350 ArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362
|||||
751 AGGCGTACCTCTATGGCAATCTGGGTGACCCACTGTCT 789
|||||
seq_name: gb_est2:B1822844

seq_documentation_block:
LOCUS B1822844 928 bp mRNA linear EST 04-OCT-2001
DEFINITION 603040166F1 NIH_MGC_115 Homo sapiens cdna clone IMAGE:5180897 5',
mRNA sequence.
ACCESSION B1822844
VERSION B1822844.1 GI:15934394
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 928)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs.rem@nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM11451 row: d column: 18
High quality sequence stop: 822.
Location/Qualifiers
source
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5180897"
/clone_11b="NIH_MGC_115"
/lab_host="DH10B"
/note="organ: pooled brain, lung, testis; Vector:
PCMV-SpOst6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dt primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH-MGC Library."
BASE COUNT 198 a 256 c 259 g 214 t 1 others
ORIGIN

```

```

alignment_scores:
Quality: 1390.50      Length: 318
Ratio: 4.762          Gaps: 5
Percent Similarity: 91.824 Percent Identity: 90.566

```

```

alignment_block:
US-09-357-675c-25 x B1822844

```

```

Align seg 1/1 to: B1822844 from: 1 to: 928
13 ProAspArgThrLeuArgMetValIleAlaIleSerSerCysArgThrTy 29
|||||
1 CCGGATCGGACCTGGCAATGCTTTTGGCTATATCTTCATGCTG..... 44
|||||
29 rSerLeuSerArgArgProArgLeuGlyPheIleThrArgProProHisA 46
|||||
45 .....GGCTTCATCACCAGCCCTCCTCACA 69
|||||
46 rGpHeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 62
|||||
70 GATTCCTGCTCCCTTCTGTCTGTCTGAGACTCCGATACCTCAACCTCAGTA 119
|||||
63 LeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCy 79
|||||
120 CTTGTGCTCAGCCCGACGCGCAGCCATGGCTATCTCTCTTCTCTCTG 169
|||||
79 sGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLysG 96
|||||
170 CGAACTGCCCTGTGCTGTGCTGTGAGGCAATGTGACTTCAATGACAGGGA 219
|||||
96 InGlnAsnPhelysThrCysAlaGluLeuValArgGluAlaIaIaArgLeu 112
|||||
220 AACAGAACTTTAAACATGTGTGAGCTGCTGAGAGGCTGCCAGACTG 269
|||||
113 GAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAs 129
|||||
270 GGTGCTGCTGCTGCTTCTGCTGAGGCAATGTGACTTCAATGACAGGGA 319
|||||
129 pProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuG 146
|||||
320 CCTGCAGAGACGCTACACCTGTCTGAAACACACGAGGGAATCTTTGG 369
|||||
146 LucIuTyrThrGlnLeuAlaArgGluCysGlyLeuThrPleuSerLeuGly 162
|||||
370 AAGAATAACACCCAGCTGGCCAGGAAATGTGACTGCTGCTCTGGTGG 419
|||||
163 GAlpHeHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 179
|||||
420 GGTTCATGAGCGGTGGCCAGACTGGGACAGACTCAGAAAATCTRCAA 469
|||||
179 nCysHisValLeuLeuAsnSerLysGlyAlaValAlaIaIaThrTyrArgL 196
|||||
470 TTGTCAAGTCTGCTGTGAGCAAGGAGGCAATGAGGCCACTTACAGGA 519
|||||
196 yStrHisLeuCysAspValAlaIuIleProGlyGlnGlyPrometCysGlu 212
|||||
520 AGACACATCTGTGTGAGTAGAGATTCCAGGAG. GGGCCTATGTGTGAA 568
|||||
213 SerAsnSerThrMetProGlyProSerLeuGluSerProValSerThrPr 229
|||||
569 AGAAGCTCTACATGCTGTGGCCAGCTTGTGAGTACACTTCAGCCACC 618
|||||
229 oAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGluL 246
|||||
619 AGAGAGCAATGTGCTGTACTGCTGTGCTATGACATCGGTTCCCTGAA 668
|||||
246 eUeSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr.TyrProse 262
|||||
669 TCTCTGTGCAATGTGCTCAAGCTGAGCAGAGATTAATTACTTATCTTC 718
|||||
262 rAlaPheGlySer...IleThrGlyProAlaHisTrrpGluValLeuLeu 278
|||||
719 AGCTTTGTGATCCCATATACAGGCGCCAGCCACTGGAGGTGTGTGCTG 768
|||||
278 rGAlaArgAlaIleGluThrGlnCysTyrVal.ValAlaIaIaIaGlnCy 294
|||||
769 GGGNCCGCTGTATCGAAACCCAGTGTATGTAGTGGCCGACACAGCTG 818
|||||
294 sGly.ArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValVal 310
|||||
819 TGAAGCGCCACCATGAGAGAGAGAGC. AGTTATGGGACACAGATGTGTGA 867
|||||

```

```

seq_name: gb_est2.B1769604

seq_documentation_block:
LOCUS      B1769604               846 bp    mRNA    linear    EST-25-SEP-2001
DEFINITION 60305498571 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204469 5',
            mRNA sequence.
ACCESSION   B1769604
VERSION     B1769604.1
KEYWORDS    GI:15761182
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 846)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/MLNLA at:
            http://image.llnl.gov
            Plate: LMAM1512 row: j column: 22
            High quality sequence stop: 838.

FEATURES
    Source
        1..846
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5204469"
            /clone_lib="NIH_MGC_122"
            /lab_host="DH10B"
            /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
            Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
            anonymous pool of 24 week female lung; 16 week female
            spleen, and 20-22 week male spleens. Library is oligo-dT
            primed and directionally cloned (pcory site is destroyed
            upon cloning). Average insert size 1.4 kb, insert size
            range 1-3 kb. Library is normalized and enriched for
            full-length clones and was constructed by C. Gruber
            (Invitrogen). Research Genetics tracking code 026. Note:
            this is a NIH_MGC Library."
BASE COUNT  179 a          240 c          226 g          201 t
ORIGIN
alignment_scores:
    Quality: 1374.00      Length: 293
    Ratio: 4.907          Gaps: 3
Percent Similarity: 95.563      Percent Identity: 95.563

alignment_block:
US-09-357-675C-25 x B1769604 ..

Align seg 1/1 to: B1769604 from: 1 to: 846

1 leuAlaProAspArgProProAspArgThrIleuArgMetValIleuAlaI1 23
|||||
3 CTGGCTCCAGACCGCCCTCCGATCGACCTCGCAATGGATTGGCTAA 52
|||||
23 eSerSerCysArgThrTyrSerLeuSerAArgArgProAArgIleuGlyPheI 40
|||||
53 ATTCTTATGCTG.....GGCTTCA 71
4 leThrArgProProIHisArgPheLeuSerLeuLeuGlySerProGlyIleuArg 56

```

[illegible]

```

ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE        1 (bases 1 to 793)
JOURNAL      NIH-MGC http://mgi.nci.nih.gov/
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: DCTP/DTP/Cazdar
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L1000382 row: 0 column: 05
              High quality sequence stop: 767.
FEATURES
  source
    1..793
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4620892"
    /tissue_type="large cell carcinoma"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: lung; Vector: pORF7; Site 1: XhoI; Site 2:
    EcoRI; cDNA made by oligo-dT priming. Directionally cloned
    into EcoRI/XhoI sites using the following 5' adaptor:
    GCCACGAG(G). Library constructed by Ling Hong in the
    Laboratory of Gerald M. Rubin (University of California,
    Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
    Superscript II RT (Life Technologies). Note: this is a
    NIH-MGC Library."
BASE COUNT   181 a 210 c 230 g 172 t
ORIGIN
alignment_scores:
  Quality: 1298.00      Length: 267
  Ratio: 4.992          Gaps: 5
  Percent Similarity: 97.378  Percent Identity: 95.880
alignment_block:
  US-09-357-675c-25 x BG436916 ..
Align seg 1/1 to: BG436916 from: 1 to: 793
81 LeupProLeuValAlaValCysGlnValThrSerThrProAspIysGlnG1 97
|||||
3 CTGCCCTGGTGGCTGTGCTGCAGGTAACTGCAGCGCCAGCAAGCAACA 52
|||||
97 naanPheLysThrCysAlaLeuValArgGluAlaAlaArgLeuGlyA 114
|||||
53 GAACCTTAAACATGCTGCTGAGCTGGTTCAGAGCGCCAGACATGGGTG 102
|||||
114 lacysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPro 130
|||||
103 CCGCGCTGGCTTCTGCTGAGGCACTTACCTTCATTCAGACGGGACCT 152
|||||
131 AlaGluThrLeuHisLeuSerGluProLeuGlyLysLeuLeuGluG1 147
|||||
153 GCAGAGACGCTACACCTGCTGAAACCACTGGGTGGAACCTTTTGAAGA 202
|||||
147 uTyrrThGlnLeuAlaArgGlyLysGlyLeuThrPleuSerLeuGlyLP 164
|||||
203 ATACACCCACCTTCCAGGAATGTGACCTGCTGCTTGGGTGCTT 252
|||||
164 heHisGluArgGlyGlnAspTrrpGluInThrGlnLysIleTyrrAsnCys 180
|||||
253 TCCATATGAGCTGGCCAAAGACTGGGAGCAGACTCAGAAAATCTCAATTGT 302
|||||
181 HisValLeuLeuAsnSerLysGlyAlaValAlaValThrTyrrArgLysTh 197

```

```

|||||
303 CACGCTGCTGAACACAGCAAGGGCAGTAGTGCCACTTAACAGAGAGAC 352
|||||
197 rHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerA 214
|||||
353 ACATCTGTGACGTAGAGATTCAGGCGAGGGCCCTATGTGTAAAGCA 402
|||||
214 snSerThrMetProGlyProSerLeuGluSerProValSerThrProAla 230
|||||
403 ACTTACCATGCCCTGGGCCAGCTTGTAGTCACCTGCACACACAGCA 452
|||||
231 GlyLysIleGlyLeuAlaValCysTyrrAspMetArgPheProGluLeu 247
|||||
453 GCGAAGATGTGCTAGCTGCTGCTATGACATGAGGGGTTCCTGAACTCTC 502
|||||
247 rLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrrPro_SerAla 263
|||||
503 TCTGGCATTTGGCTCAACGTGGAGCAGAGATTACTATCTTCACAGCT 552
|||||
264 PheGlySerIleThrGlyProAlaHisTrrpGluValLeuLeuArgAla_A 280
|||||
553 TTTCGATCCATTCACAGGCCACAGCCACTGGAGAGTGTTCCTGGCGGCC 602
|||||
280 rGlaIaIleGluThrGlnCysTyrrValAlaAlaAlaAlaGlnCysGlyArg 296
|||||
603 GTGCTATGCAAAACCCACTGCTATGATGACAGCAGCAGCATGTGCAGCC 652
|||||
297 HisHisLysArgAlaSerTyrrGlyHisSerMetValAlaAsp_ProT 313
|||||
653 CACCATGAGAGAGAGCAAGTTATGGCCACAGCATGTGTGTAGACCCCT 702
|||||
313 rPgly_ThrValAlaAlaArgCysSer_GluGlyProGlyLeuCysLeuA 329
|||||
703 GGGGAACAGTGGTGGCCCTGCTGTTAGGGGCCAGGCTTCGCTTGC 752
|||||
329 laarGllaasPleuAsnTyrrLeuArgGlnLeuArgArgHis 342
|||||
753 CCCGAATAGACATCAAGTATCTGCGACAGTGGCCGCAAAAC 793

```

```

seq_name: gb_est2:BG703176
seq_documentation_block:
  LOCUS      BG703176              790 bp      mRNA      linear      EST 07-MAY-2001
  DEFINITION  602686082F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4818405 5',
  mRNA sequence.
  ACCESSION  BG703176
  VERSION    BG703176.1 GI:13975252
  KEYWORDS   EST.
  SOURCE      human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
              1 (bases 1 to 790)
  REFERENCE  NIH-MGC http://mgi.nci.nih.gov/
  AUTHORS    National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE      Unpublished (1999)
  JOURNAL    Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
              cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
              Toshiyuki and Piero Carninci (RIKEN)
              DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L10010720 row: d column: 22
              High quality sequence stop: 766.
FEATURES
  source
    1..790
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4818405"

```


Percent Similarity: 93.985 Percent Identity: 93.985

alignment_block:

US-09-357-675C-25 x BG762506 ..

Align seg 1/1 to: BG762506 from: 1 to: 764

```

10 AsparGProProAsparGThrLeuArGMeValLeuAlaIleSerGercy 26
2 GACCGCCCTCCGATCGACCGTGCAGATGGTTTGGCTATATCTTCAC 51
26 sargThrTyrSerLeuSerArGArProArGLeuGlyPheIleThrArp 43
52 CCG.....GGCTTCATCCAGCG 70
43 roProHisArGPhelSerLeuLeuGlySerProGlyLeuArGLeuProGln 59
71 CTCCTACAGATTCTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 120
60 .LeuSerValLeuGlyAlaGlnProArGProArGAlaMetAlaIleSer 76
121 ACTCTGACTGTTGTGTGCTACGCCAGGCCAGAGCCATGCTATCTCT 170
76 eSerSerGlyGluLeuProLeuValAlaValCysGlnValThrSerThr 92
171 CTTCCTCCGCGAAGTCCCTGCTGCTGTGTGCGAGGTACATCATGACG 220
93 ProAspGlnGlnAspPheLysThrCysAlaGlnLeuValArgGlu 109
221 CCAGACACAGACAGACACTTTAAACATGCTGCTGCTGCTGCTGCTGCT 270
109 aAlaArGLeuGlyAlaCysLeuAlaPheLeuProGlnAlaPheAspPhe 126
271 TGGCAGACAGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 320
126 lAlaArGAspProAlaGluThrLeuHisLeuSerGluProLeuGly 142
321 TTGCAGGGGACCTCGACAGACCTACCTGCTGCTGCTGCTGCTGCTGCT 370
143 LysLeuLeuGluGluThrThrGlnLeuAlaArgGluCysGlySerPhe 159
371 AACCTTGGGAGAGATACACCCAGCTTCCAGGGAGATGAGACTGTGGCT 420
159 uSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 176
421 GTCCTTGGGTGGTTTCCATGAGCGGTGCCAAGACTGGAGACAGTCA 470
176 ySleIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 192
471 AAATCTACAAATTGTCACTGCTGCTGAACAGCAAGGGGAGAGTGGCC 520
193 ThrTyrArGlySerHisLeuCysAspValGlnIleProGlyGlnGly 209
521 ACTTACAGAGACACATCTGTGTGACGTAGAGATTCCACGGCAGGGGCC 570
209 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 226
571 TATGTGTGAAGCACTTACCATGCTGGGCCCACTTGACTCACTG 620
226 aLSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetAr 242
621 TCAGCAGACAGCAGCAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670
243 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluLeuThr 259
671 TTCCCTGAACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
259 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrp 273
721 CATCTCTTCAAGCTTTTGATTCATTACAGGCCAGGCCACCTGG 764

```

seq_name: gb_esl1:AL559163

seq_documentation_block:

LOCUS AL559163 861 bp mRNA linear EST 16-FEB-2001

DEFINITION LTI_NFL008_TC2 Homo sapiens cDNA clone CS0DJ014YA17 5 prime, mRNA sequence.

ACCESSION AL559163

VERSION AL559163.1 GI:12904391

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 861)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..861

location/Qualifiers

organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DJ014YA17"

/clone_1lb="LTI_NFL008_TC2"

/sex="male"

/tissue-type="T cells from T cell leukemia"

/note="Vector: pcWSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcWSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@life.com URL : http://fulllength.invitrogen.com"

BASE COUNT 176 a 229 c 225 g 201 t 30 others

alignment_scores:

Quality: 1263.00 Length: 299

Ratio: 4.730 Gaps: 5

Percent Similarity: 89.298 Percent Identity: 86.288

alignment_block:

US-09-357-675C-25 x AL559163 ..

Align seg 1/1 to: AL559163 from: 1 to: 861

```

17 LeuArMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerAr 33
2 CTCKGATAGGTTTGGCTATATCTTCATGCTG..... 33
33 gArGProArGLeuGlyPheIleThrArGProProHisArGPhelSerL 50
34 .....GGCGTCAATCACAGGCGCTCTCAGACTTCCTGCTGCC 70
50 euleuGlySerProGlyLeuArGLeuArgIleProGlnLeuSerValLeuGlyAlaGln 66
71 TTCTGTGTCCTGACATCCGATACCTCAACTCTCACTACTTGTGCTCAG 120
67 ProArGProArGAlaMetAlaIleSerSerSerCysGluLeuProle 83
121 CCCAGGCCAGAGCCATGCTATCTCTCTCTCTGCGAAGTGGCCCT 170
83 uValAlaValCysGlnValThrSerThrProAspLysGlnGlnAspPhe 100
171 GGGGGCTGTGTGCGACGTAACATCGACGCCAGATYAGCAATGACTTTA 220
100 yThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeu 116

```

```

221 AACATGTCGTGAGCTGTTGAGAGGCTGCCAGACTGGTGCCCTGCCCTG 270
117 AlpheleupProgluaIaPheaspPheIIaIaArgasProAlaIgluTh 133
|||||
271 GCTTTCCTGCTGAGGATTTGACTTCATTCAGTGGAGACCCTGCAGAGAC 320
133 rIeuHLSleuSerGluPro-leuGlyGlyLeuLeuGluGluThrThr 149
|||||
321 GCTACATCTCTGACCTACTGGGTGGAAACTTTGGAAAGATACATC 370
150 GlnLeuAlaArgGlu-CysGlyLeuTrpLeuSerLeuGlyGlyPheHisG 166
|||||
371 CACGTTCCACAGGATATGTGACCTGCTGCTGCTGCTGGGTTCATG 420
166 LuArgGlyGlnAspTrpGluGlnThrGlnLysIleTyraAsnCysHisVal 182
|||||
421 AGCGTGCGYMWGACTGGGAGACACTYAGAAACTCAATTTGCACGTG 470
183 LeuLeuAsnSerLysGlyAlaValAlaIaIaThrTyraArgLysThrHisI 199
|||||
471 CTGCTGACACGCAAAAGGCGAGTAGTGCCACTTACAGAGAACACATCT 520
199 uCysAspValaGluIleProGlyGlnGlyProMetCysGluSerAsnSer 216
|||||
521 GTGAGAGTAGAGATTCCAGGCGAGGGCCCTATGTGAAAGCAACTCTA 570
216 hrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLys 232
|||||
571 CCATGCGTGGGCCACGCTTGAGTACACCTGTACAGACACACGAGCAAG 620
233 ILleGlyLeuAlaValCysTyraSpMetArgPheProGluLeuSerLeuAl 249
|||||
621 ATTGCTCTACCTGCTGTGTATGACATGCGTTCCTGCACCTCTCTC 670
249 AluAlaGlaGlaGlyAlaGluIleLeuThrTyProSerAlaPheGlyS 266
|||||
671 ATTGGCTCMAAGCTGGCCAGAGACTTMACTTMYCTTCMCCTTTGGCW 720
266 erile-ThrGlyProAlaHisTrp-GluValLeuLeuArgAlaArgAlaI 282
|||||
721 CCATTACAGGCCACGCCACTGGGAGGTGTCTGCGGGGCCGCTGCTC 770
282 leguThrGlnCysTyraValAlaAlaAlaGlnCysGlyArgHisHis 298
|||||
771 WCGMAAACCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 820
299 GluLysArgAlaSerTyraGlyHisSerMetValValasp 311
|||||
821 GAGAGAGAGCAACTTATGGCCACAGCATGTGTGTAGAC 859

```

```

seq_name: gb_est2:BI752623
seq_documentation_block:
LOCUS      BI752623                746 bp    mRNA    linear    EST 25-SEP-2001
DEFINITION 603028471ef1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5198800 5',
ACCESSION  BI752623
VERSION     BI752623.1 GI:15744201
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 746)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            JOURNAL
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLML1497 row: n column: 17

High quality sequence stop: 746.

Location/Qualifiers

1..746

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NIH_MGC_114"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;

Site:2: EcoRV (destroyed); RNA source anonymous pool of 6

male brains, age range 23-27 yo. Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.5 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 019. Note:

this is a NIH_MGC Library."

```

BASE COUNT      154 a      220 c      197 g      175 t
ORIGIN

```

```

alignment_scores:
  Quality: 1237.50      Length: 249
  Ratio: 5.135          Gaps: 3
  Percent Similarity: 96.787      Percent Identity: 96.787

```

alignment_block:

US-09-357-675C-25 x BI752623 ..

Align seg 1/1 to: BI752623 from: 1 to: 746

```

7 LeuAlaProAspArgProProAspArgThrLeuAlaGlyMetValLeuAlaI 23
|||||
4 CTGGCTCCAGACCGCCCTCCGAGTGGACCCCTGCAATGATTTGGCTAT 53
23 eSerSerCysArgThrTyraSerLeuSerArgArg.....P 35
|||||
54 ATCTTCATGTAGGAGACTACTCCCTATCCCTGCGGGCGGATATCCAC 103
35 ro.ArgLeuGlyPheIleThrArgProProHisArgPheLeuSerLeu 51
|||||
104 CTGGGCGTGGGCTTCATCACAGGCTCTCTCACAGATTCTCTCCCTCT 153
51 uCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro 68
|||||
154 GTGTCTGAGACTCCGATTAACCTCAACTCTCACTACTTGTGTCAAGCCA 203
68 rGProArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuVal 84
|||||
204 GGGCCAGAGCCATGAGCTATTCCTCTCTCTCTCTGGAAGTGGCCCTGT 253
85 AlalaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhaLys 101
|||||
254 GCTGTGTCACAGGTAACTACACGCCAGCAAGCAAGCAAGCAAGCAAG 303
101 rCysAlaGluLeuValaArgGluAlaAlaArgLeuValaCysLeuAlaP 118
|||||
304 ATGTGCTAGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTGGCTGG 353
118 hLeuPProGluAlaPheaspPheIleAlaArgasProAlaGluThrLeu 134
|||||
354 TCCTGCTGAGGCAATTCATTCATTCAGAGGAGCCATGCAAGAGCGTA 403
135 HisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluGluThrGln 151
|||||
404 CACCGTCTGAACCCACTGGGTGGAAACTTTGGAAGATACACCCACT 453
151 uAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluVal 168
|||||

```

454 TGGCAGGAAATGTGACTGTGCTGTCTGGTGGTTCCTCCAGACGCTG 503
168 lylgnasptpgluglnthrglnlysllethyrasnrcysHsvalleu 184
|||||
504 GCCAGACTGGAGACACTCAGAAATCTACAACTTCACGTCGCTGCTG 553
185 AsnSerIysGlyAlaValAlaIleThrTyrArgLysThrHisLeuCys 201
|||||
554 AACAGCAAGAGGGGACAGTGTGGCCACTTACAGAGAAACACATCTGTGA 603
201 pvalgluileproglnglynglyprometCysgluserAsnSerThrMet 218
|||||
604 CGTAGAGATTCCAGGCGAGGCGCTATGTGTGAAGCAACTCAACATGC 653
218 roglproserleugluserProvalSerThrProalGlyLysIleGly 234
|||||
654 CTGGGCCCACTCTTGAGTCACTGTGACAGACACAGCAGGCAAGATAGT 703
235 leuAlaValCysTyrAspMetArg.PheProgluLeuSerleu 248
|||||
704 CTAGCTGTCTGTATGACATGCGGGTTCCTGAACTCTCTGTG 746
seq_name: gb_est2:BI755493

seq_documentation_block:
LOCUS BI755493 813 bp mRNA linear EST 25-SEP-2001
DEFINITION 603027349P1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5197716 5',
mRNA sequence.
ACCESSION BI755493
VERSION BI755493.1 GI:15747071
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 813)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Place: LLM11495 row: a column: 13
High quality sequence stop: 745.

FEATURES
source
1..813
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5197716"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH-MGC Library."
BASE COUNT 155 a 241 c 225 g 192 t
ORIGIN

alignment_scores:
Quality: 1223.50 Length: 278
Ratio: 4.817 Gaps: 3

Percent Similarity: 91.367 Percent Identity: 89.209

alignment_block:
US-09-357-675c-25 x BI755493 ..

Align seg 1/1 to: BI755493 from: 1 to: 813

1 ProleuAlaAlaAla**LeuAlaProAspArgProProAspArgThrle 17
|||||
13 CCATGCTCGGGGCTTCTGTGGCTCCAGACCGCCCTCGGATGGACCT 62
17 uArgMetValLeuAlaIleSerSerCysArgThrTyrSerLeuSerArg 34
|||||
63 GCGAATGTTTGGCTATATCTTCATGCTG..... 92
34 rgrProArgleuGlyPheIleThrArgProProHisArgPheLeuSerleu 50
|||||
93GGCTTCATCACCGAGCCCTCTCCACAAATTCCTGCTT 131
51 LeuCysProglYleuArgIleProglInLeuSerValLeuCysAlaGlnpr 67
|||||
132 CTGTGCTCTGAGACTCCGATACCTCAACTCTCACTACTTGTGCTCAGCC 181
67 oArgProArgAlaMetAlaIleSerSerSerSerSerCysgluLeuProleu 84
|||||
182 CAGGCCAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 231
84 AlAlaValCysGlnValThrSerThrProAspLysGlnGlnAspPheLys 100
|||||
232 TGCGTGTGTGCCAGTAACTCAACGCCAGACCAAGCAAGCAACTTTAA 281
101 ThrCysAlaGluLeuValArgGluAlaAlaArgleuGlyAlaCysLeuAl 117
|||||
282 ACATGCTCTGAGCTGTTTCAGAGAGCTGCCAGACTGGGTCCTGCTGCG 331
117 apheLeuProgluAlaIleAspPheIleAlaArgAspProAlaGluThr 134
|||||
332 TTTCTCTCTCTGAGGCAATTCATTCATTCAGACGGACCTGCAGAGAGCG 381
134 euHisLeuSerGluProleuGlyGlyLysLeuLeuGlnGluTyrThrGln 150
|||||
382 TACACCTGTGTGAACCACTGGTGAGAACTTTTGGAGAAATACACCCAG 431
151 leuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluAr 167
|||||
432 CTTCGCGAGGAATGTGACTGTGCTGCTGCTGGTGGTTCATGAGCG 481
167 gGlyGlnAspTrpGluGlnThrglnlysllethyrasnrcysHsvalleu 184
|||||
482 TGGCCAAAGACTGGAGCAGCAGACTCAGAAATCTTACAAATTCACGCTGTC 531
184 euAsnSerIysGlyAlaValAlaIleThrTyrArgLysThrHisLeuCys 200
|||||
532 TGAACACACCGGGGCGAGTGTGGCCACTTACAGAAACACATCTGTGT 581
201 AspValGluIleProglnglynglyPrometCysgluserAsnSerThrMe 217
|||||
582 GAGTAGAGATTCAGGCGAGGCGCTATGTGTGCAAGCAACTTACCAT 631
217 tProglYProserleugluserProvalSerThrProalGlyLysIleG 234
|||||
632 GCCTGGGCCCAAGCTTGAAGTACCTGTACAGCGAGCGAGCGAG..ATTTG 680
234 lylleuAlaValCysTyrAspMetArgPheProgluLeuSerleuAlaLeu 250
|||||
681 GGCTAGCTGTCTATGAGCTGGGGTTCCTGAACCTCTCTGGGCAATTTG 730
251 Alagln..AlaGlyAlaGluIleLeuThrTyrProSerAlaPhe..... 264
|||||
731 GCTCAGACGCTGGAGCGGAGATACCTTACCTATCTTACGCTTGGCGTCT 780
265GlySerIleThrGlyPro 270
|||||

781 TGCGGGCCCGCACCTGAGAGCGGTGCGGGGCT 813
seq_name: gb_est2:BI757823
seq_documentation_block:
LOCUS BI757823 946 bp mRNA linear EST 25-SEP-2001
DEFINITION 603030404P1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200560 5',
mRNA sequence.
ACCESSION BI757823
VERSION BI757823.1 GI:15749401
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 946)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
DNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11502 row: h column: 01
High quality sequence stop: 708.
Location/Qualifiers
FEATURES
source 1..946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5200560"
/clone_id="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH-MGC library."
BASE COUNT 213 a 264 c 270 g 199 t
ORIGIN
alignment_scores:
Quality: 1219.00 Length: 324
Ratio: 4.247 Gaps: 8
Percent Similarity: 88.580 Percent Identity: 81.790
alignment_block:
US-09-357-675c-25 x BI757823 ..
Align seg 1/1 to: BI757823 from: 1 to: 946
3 AAlaAlaAla**leuAlaProAspArgProAspArgThrIleuArgme 19
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
2 GCTGCGCGCTTCTGCTCGACACCGCCCTCCGATGCGACCCGCGAAT 51
||| tValleuAlaIleSerSerCysArgThrTySerIleuSerArgProA 36
1 | ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
52 GGT.TTGGCTATATCTTCATGCTG..... 74
36 rgleuGlyPheIleThrArgProProHisArgPheIleuSerIleuCys 52
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
75GGCTTCATCACCGAGCGCTCTCACAGATTCCTGCTCTGTGT 119
53 ProGlyIleuArgIleProGlnIleuSerValIleuCysAlaGlnProArgpr 69

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
120 CCTGACTCCGGATACCTCAACTCTCAGTACTTGTGCTCAGCCAGGCC 169
69 oArgAlaMetaAlaIleSerSerSerSerCysGluLeuProIleuValAla 86
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
170 CAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 219
86 aLysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 102
220 TTGTCCAGGTAAACATTCAGACCGACAGACAAACGAACTTTAAACATGT 269
103 AlaGluIleuValArgGluAlaAlaArgIleuGlyIleuAlaPheLe 119
270 GCTAGCTGTTGCGAGAGCTGCCAGACTGGGTGCTGCTGCTGCTTCT 319
119 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrIleuAs 136
320 GCCTGAGGCAATTGACTTCATTGACGCGGACCGCTGACAGACGCTACAC 369
136 euSerGluProIleuGlyLysIleuLeuGluGluTyThrGlnLeuAla 152
370 TGTCTGAACACACTGGGTGGGAACTTTGGAAAGATACCCAGCTTGGC 419
153 ArgGluCysGlyLeuThrIleuSerIleuGlyIleuHisGluArgGly 169
420 AGGAAATGTGACTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 469
169 nAspTrpGluGlnThrGlnLysIleTyAsnCysHisValIleuLeuAs 186
470 AGACTGGGAGAGAGACTCAGAAATCTCAAGTGCACGTGCTGTGTAACA 519
186 eLysGlyAlaValAlaThrTyArgLysThrHisLeuCysAspVal 202
520 GCAAGGCGGCGAGTAGTGCCACTTACAGAGACACATCTGTGACACTA 569
203 GluIleProGlyGlnGlyProMetCysGluSerIleuSerIleuMetPro 219
570 GAGATTCAGGCGAC. GGGCTATGTGTGAAGCAACTCTACCATGCTGG 618
219 yProSerIleuGluSerProValSerThrProAlaGlyLysIleuLeu 236
619 GCCAGAGCTTGAAGTACCTGTCACACACCGGAGGCAAGTTGTGTAG 668
236 IavaLysTyArgPheMetArgPheProGluLeuSerIleuAlaLeuAla 252
669 CTGCTCTATGACATGCGGTTCCTGAATTCCTCGGCGAGTGCATCA 718
253 AlaGlyIleuGluIleuThr.TyrProSerAlaPheGly.SerIleThr 268
719 GCTGAGCAGAGATTAATTAATCCGTACGCTTGGATCAGATTACG 768
269 GlyPro.....AlaHisTrpGluValIleuLeuArgAlaAlaI 282
769 GGGCCAGACACCTGGAGAGGTGAAG.....CTGAGGGCCAGTGGCT 812
282 e...GluThr...GlnCysTyArgValAlaAlaAlaGlnCysGlyArg 297
813 ATCCGAAACGCTAGAGGCTCAGATATGTCGCGACAGCACACGCG. 861
297 is...HisGluLysArgAlaSerTyArgIleHisSerMetValValAs 312
862 ACCCTGACAGAGAGAGCCAGATTAAGGCGCACAGATGTGTGAAAAACC 911
313 TrpGlyThrValAlaAla 318
912 ATGGGAAACAGAGATGCC 929
seq_name: gb_est1:AL520767
seq_documentation_block:
LOCUS AL520767 992 bp mRNA linear EST 13-FEB-2001
DEFINITION AL520767 LFI_NF1004 NBC2 Homo sapiens cDNA clone CS0DB002YD05 3
prime, mRNA sequence.

ACCESSION AL520767
 VERSION AL520767.1 GI:12784260
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 992)
 Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1..992
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DB002YD05"
 /clone_id="LTI_NFL004_NBC2"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(dt) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life Technologies. Contact : Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
 8371 Email : fliang@lifetechn.com URL :
 http://fulllength.invitrogen.com"
 BASE COUNT 229 a 267 c 268 g 222 t 6 others
 ORIGIN
 Alignment_scores:
 Quality: 1201.50 Length: 236
 Ratio: 5.179 Gaps: 3
 Percent Similarity: 98.305 Percent Identity: 96.610
 Alignment_block:
 US-09-357-675C-25 x AL520767/rev ..
 Align seg 1/1 to reverse of: AL520767 from: 1 to: 992
 129 AspproAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuLeu 145
 |||||
 983 GACCTGCAGACGCTACCTGCTGAAACCACTGGTGAAAC...TT 937
 |||||
 145 ucluglu TyrthrGlnLeuAlaArgGluGlyLeuTrpLeuSerIeu 161
 : : : : :
 936 TtGRRAgRAATACACCCAGCTTGCAGGAAATGGACTGTGCTCTTG 887
 |||||
 162 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnIleTyr 178
 |||||
 886 GGTGGTTCATGAGAGCTGGCAGACTGGGAGCAGACACAGAAATCTA 837
 |||||
 176 tAsnCySHisValLeuLeuAsnSerIysGlyAlaValAlaIaThrTyr 195
 |||||
 836 CAATGTCTACGCTGCTGTAAACAGCAAGGGGCGAGTAGTGGCCACTTACA 787
 |||||
 195 tGlyThrHisLeuGlyAspValGluIleProGlyGlnGlyProMetCys 211
 |||||
 786 GGAAGACACATCTGTGTAGCTAGAGATTCCAGGGGAGGGGCTATGTGT 737
 |||||
 212 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 228
 |||||
 736 GAAAGCAATCTACCATGCTGGGCCCATCTTGAATCACTGTCTGACGAC 687
 |||||
 228 tProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPhePro 245

|||||
 686 ACCAGCAGCAGCAAGATTGGTCTACTCTGTCTATGACATGCGGTCCG 637
 |||||
 245 ludeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 261
 |||||
 636 MACTCTCTGTGCATTTGGCTCAAGCTGGAGCAGACAGATCTACTTACTCT 587
 |||||
 262 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 278
 |||||
 586 TCAGCTTTTGATCCATTACAGGCCAGCCCACTGGAGAGTGTGGCTGGCG 537
 |||||
 278 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCys 294
 |||||
 536 GGGCCGTGTATCGAAACCCAGAGCTATGTAGTGCCTCCAGCAGCAGTGT 487
 |||||
 295 GlyArgHisHisGluAsnArgAlaSerTyrGlyHisSerMetValVal 311
 |||||
 486 GAGAGCCACCATGAGAGAGAGCAACTTATGGCCACAGATGGTGTATA 437
 |||||
 311 pProTrpGlyThrValAlaAlaArgCysSerGluGlyProGlyLeuCysAl 328
 |||||
 436 CCCCTGGGGAACAGTGTGGCCGCTGTCTGAGGGGCCAGGCTGTGCC 387
 |||||
 328 euAlaArgIleAsnLeuAsnTyrLeuArgGlnLeuArgHisLeuPro 344
 |||||
 386 TTGCCGGAATAGACCTCACTATCTGCAGAGTGGCCGACACCTGCT 337
 |||||
 345 ValPheGlnHisArgArgProAsnLeuTyrGlyAsnLeuGlyHisProle 361
 |||||
 336 GTTCTTCAGACCGCAGGCTGACCTTATGGCAATCTGGTCCACT 287
 |||||
 361 uSer 362
 |||||
 286 GTCCT 283
 seq_name: gb_est1:AL522373
 seq_documentation_block:
 LOCUS AL522373 960 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL522373 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB008YM02 3
 prime, mRNA sequence.
 ACCESSION AL522373
 VERSION AL522373.1 GI:12785866
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 960)
 Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1..960
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DB008YM02"
 /clone_id="LTI_NFL004_NBC2"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(dt) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life Technologies. Contact : Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610

8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 225 a 259 c 255 g 217 t 4 others
ORIGIN

alignment_scores:
Quality: 1179.50 Length: 231
Ratio: 5.151 Gaps: 2
Percent Similarity: 99.134 Percent Identity: 98.268

alignment_block:
us-09-357-675c-25 x AL522373/rev ..

Align seg 1/1 to reverse of: AL522373 from: 1 to: 960

```
133 ThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeu.GluGluTyrT 149
|||||
959 ACCGTAACACCTGCTGTGAACCACTGGGTGGAAC...TTTGAGAAATACA 913
149 hrcLleuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHis 165
|||||
912 CCCAGCTTGGCCAGGAAATGTGGACTGTGCTTGGGTGGTTCAT 863
166 GluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisVa 182
|||||
862 GACCGTGCCCAAGACTGGAGCAGACTCAGAAATCTACAAATTGTCAGT 813
182 lleuLeuAsnSerLysGlyAlaValAlaIleThrTyrArgLysThrHisL 199
|||||
812 GCTGTGAACAGCAAGGGGCGAGTGGCCACTTACAGGAGACACATC 763
199 euCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 215
|||||
762 TGTGTGACGTAGAGATTCCAGGGCAGGGGCTTGTGTGAAGCAACTCT 713
216 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLy 232
|||||
712 ACCATGCTGTGGCCAGTCTTGTGAGTACCTGTCAGCAGACACAGAGCAA 663
232 sileGlyLeuAlaValAlaCysTyrAspMetArgPheProGluLeuSerLeuA 249
|||||
662 GATTGTGTAGTGTGTGTGATGACATGGC.TTCCGTGAACCTCTCTGG 614
249 laLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 265
|||||
613 CATTGGCTCAAGCTGGAGCAGAGATACTACTATCCTCAGCTTTTGA 564
266 SerIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaI 282
|||||
563 TCCATTACAGGCCAGCCAGCTGGAGGTGTGTGGGCCCGCTGTAT 514
282 egluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisG 299
|||||
513 CGAAACCCAGTGTATGTAGTGGCAGCAGCAGAGTGTGGAGCCACCATG 464
299 luLysArgAlaSerTyrGlyHisSerMetValAlaAspProTrpGlyThr 315
|||||
463 AGAAGAGAGCAAGTATGGCCACAGCATGTGTGTAGACCCCTGGGGAACA 414
316 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAs 332
|||||
413 GTGGTGCCCGGTGCTCTGAGGGGCCAGGCTTGCCTTGGCCGAAATAGA 364
332 PleuAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisA 349
|||||
363 CTCACACTATCTCGACAGATTGGCCGACACACTGCTGTGTCACAGCACC 314
349 rgaArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 362
|||||
313 GCAGGCCTGACCTCTATGCAATCTGGGTGACCCACTGTCT 273
```

THIS PAGE BLANK (USPTO)

OM of: US-09-357-675C-25 to: Issued_Patents_NA: * out_format : pfs
Date: Apr 29, 2002 10:03 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed.p2n.model -DEV=xlh
-O/cgn2_1/USPFO.spool/US09357675/runt_29042002_091757_3310/app_query.fasta.1.1208
-DB=Issued_Patents_NA -QPM=fastaf -SUFFIX=p2n.rml
-GADP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -OGADP=4.500 -OGAPEXT=0.050 -XGADP=10.000
-XGAPEXT=0.500 -FGADP=6.000 -FGAPEXT=7.000 -YGADP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=dlsm02 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09357675 @CGN1.118 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-357-675C-25
Query length: 362
Database: Issued_Patents_NA: *
Database sequences: 383533
Search length: 122816752
Search time (sec): 85.140000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63 +		215.50	363.13	7.1e-13	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-63 +		215.50	363.13	7.1e-13	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-25 +		214.50	367.25	9.1e-13	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-25 +		214.50	367.25	9.1e-13	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-5 +		213.50	365.38	1.2e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-5 +		213.50	365.38	1.2e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-5 +		213.50	365.38	1.2e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-5 +		213.50	365.38	1.2e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-17 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-17 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-19 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-19 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-61 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-61 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-17 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-17 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-19 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-19 +		212.50	363.50	1.5e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-9 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-9 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-21 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-21 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-9 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-9 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-11 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-11 +		211.50	361.63	1.9e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-917-111-1 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-07-917-111-1 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-1 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-1 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-3 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-3 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-7 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-7 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-15 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-15 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-23 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-23 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-49 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-49 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-51 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-51 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-55 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-55 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-65 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-294-871A-65 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-1 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-1 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-3 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-3 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-876-398A-7 +		210.50	359.75	2.4e-12	1785	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-7 +		210.50	359.75	2.4e-12	1785	

/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-15 + 210.50 359.75 2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-23 + 210.50 359.75 2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-27 + 210.50 359.75 2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-29 + 210.50 359.75 2.4e-12 1785
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-876-398A-31 + 210.50 359.75 2.4e-12 1785

seq_name: /cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-294-871A-63

seq_documentation_block:

Sequence 63, Application US/08294871A
Patent No. 5824522
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiratsuki, Yoshio
TITLE OF INVENTION: Immobilized Enzyme Preparation and
PROCESS OF INVENTION: Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294, 871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP92/00739
FILING DATE: 10-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01101
FILING DATE: 05-AUG-1993


```

APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 340078/1992
FILING DATE: 21-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 212692/1992
FILING DATE: 10-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wegner, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 74129/130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEXFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 1785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: JM109 PAD469
FEATURE:
NAME/KEY: CDS
LOCATION: join(233..1141)
US-08-876-398A-63
alignment_scores:
Quality: 215.50 Length: 406
Ratio: 1.094 Gaps: 17
Percent Similarity: 48.522 Percent Identity: 24.384
alignment_block:
US-09-357-625 x US-08-876-398A-63 ..
Align seg 1/1 to: US-08-876-398A-63 from: 1 to: 1785
2 LeuAlaIaIaIa***LeuAlaProAspArgProProAspArgThIeuAr 18
|||||:|||||||::: ||| |||||:::
26 TTGCAAGACGCCAAATTCGGGTCCGCTCGCGTTGACAGATCAAATAAF 75
18 gmeValLeuAlaIeSerCyS..... 26
::: ||| ::||| |||
76 TTTAAGCGCTGTACTGCATCATATATTTCTACTTATAGTAG 125

```

27ArghTlYrSerLeuSerAlaGdApProAlaGleUglyPhe 39
126 AATTGCATTGGCGCGGAGTCCAAAGCCGGTTTTCGGCGATGTGTTTC 175
40 IleThr.....ArgProProHisat 46
176 ACAACGTTTCCCGCGCGGTGGCGGAGACATACCCTAGGAAGAGCAGAG 225
46 gphLeuSerLeuUcysProGlyLeuAlaArgIleProGlnLeuSerVal 63
226 GTTCATGAGACA.....CGTCAGATGATCTATTG 251
63 eucSalalagInProAlrProAlrGalaMetAlaIleSerSerSerCys 79
252 CAGTGGGACACAAAGGTCGATCGCGCGCGGAGACACGGGAACAGCTC 301
80 GluLeuProLeuValAlaValAlcLysGlnValThrSerThrProAspLysG 96
302 GTCGTCGCTCTTCGCACATG.....CTGACGAAGCGCGGACCGGG 345
96 ngInAsnPhelYThrCysAlaGluLeuValAlaGluValAlaAlaArgLeuG 113
346 CGGAGATTTCATTGCTGCTCCCGGACATC..... 373
113 LylalacysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgP 129
374 ..CGCTTACAGACCTTCTCCCGCGC...TGCGATTCTACCGAGAGGCC 418
130 ProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuG 146
419 GAGCTCGATAGCTCTCATAGACCGAA...ATGCGCGCGCGGATGGTCCG 465
146 uGluTYrThrGlnLeuAlaArgGluCysGlyLeuThrPheSerLeuGlyG 163
466 TCCACCTCTTGAGAGAGCGCGGGAATCGGATCGCTCAATCTGGGCT 515
163 LyrPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnLys 176
516 ACCTGACGATCGTCGTGAGAGCGGCG.....GTCAACGCT 550
177 IleTYrAsnCysHisAlaLeuLeuAsnSerLysGlyAlaValValAlaThr 193
551 CGCTCAACACGTCCTATTGGTGGATGAATGACGACCAAGCATCGTCGCA 600
193 rTYrArgLysThrHisLeuCysAspValGluIleProGlyGlnLysProm 210
601 GTATCGTAGAATCCATTTG.....CGGGTCACAAAG... 631
210 eTcysGluSerAsnSerThrMetProGlyProSerLeuGluSerProVal 226
632GAGTACGAGCGCTACCGCGCGTTCACAGACATCTTGAAACCGTTAT 276
227 SerThrPro.....AlaGlyLys 232
677 TTCCAGCGCGGCGGANTCGGCTTCCGGCTATATGACGTGACGCGCGGAA 726
232 sIleGlyLeuAlaValCysTYrAspMetArgPheProGlyLeuSerLeuAla 249
727 AATGGGAGATTGATGTGCACAGATCGCGCGCTGGCTTAACCTGGCGGG 776
249 lAlaValaGlnAlaGlyAlaGlyIleLeuThrTYrProSerAlaPheGly 265
777 TGATGGCGCTCAGAGGCGCGCGAGATCATGTGC.....GGC 811
266 SerIleThrGlyProAlaHisTrpAlaValLeu..... 276
812 GCGTACACACGCGCGACACCAATCCCATGTGTCCCGACGACGACACT 861
277LeuArgLalaArgLalleGluThrG 285
862 GAGCTCTTCCACCATCTCTCATGATGACGAGCGCGGCTTATACAGACG 911
285 lncSTYrValValAlaAlaAlaAlaGlnCysGlyArgHisHisGluLysArg 301


```

302 GTCGTCGTCGTCGACATG.....CTGACGAAGCCGCGAGCCGGG 345
96 nglnasphelythrCysalagluValargluAlaIaIargLeug 113
346 CGCGAATTTCATTGTCTCCCGAATC..... 373
113 lYalacysLeuAlaPheLeuProgluAlaPheAspPheIlaIaIargsp 129
374 ..GGCCTTAGACCTTCTCCCGGCC..TGGCATTTCCAGCAGAGGCC 418
130 ProAlaIuThrLeuHisLeuSerGluProLeuGlyGlyLeuLeuG1 146
419 GAGCTGATAGCTTCTATGAGACGAA...ATGCCCGCCGCGGTGTCG 465
146 uGluTyrThrGluLeuAlaIargGlyGlyLeuTyrLeuSerLeuGly 163
466 TCCACTCTTGAGAGCGCGGAACCTCGGATCGCTCAATCTGGGCT 515
163 lYpHeHis.....GluArgGlyGlnAspTyrGluThrGlnLys 176
516 ACGCTGAACTGCTGTCGAGAGCGGC.....GTCAAGCGT 550
177 lLetYrAsnCysHisValLeuLeuAsnSerGlyAlaValAlaTh 193
551 CGCTCAACAGCTCCATTGTTGGATAGTCAGCAAGATCGTCGCCAA 600
193 rTYrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProm 210
601 GTATCGTAAGATCCATTGTG.....CCGGGTCCACAAG... 631
210 etCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProVal 226
632 ....GAGTACGAGGCGCTACCGCGCTTCCAGCATTTGAAGAGCTTAT 676
227 SerThrPro.....AlaGlyLys 232
677 TTTCAGCGCGCGCATCTCGGCTTCCCGCTATAGTCGACGCGCGCAA 726
232 sILeGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuA 249
727 AATGGGATGTTTCATGTCAGACGATCGCGCTGAGTCAAGCTGCGCGG 776
249 lAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr.....TyrProSerAla 263
777 TGATGGGCTTCAGGGCGCGGATCATCTCGGCGCTACACACGCGG 826
264 PheGlySerIleThrGlyProAlaHisTyrGluValLeu..... 276
827 ACCCAATCCACCGCTTCCCGACAGCACACCTGACGTCCTTCCACCA 876
277 .....LeuArgAlaArgAlaIleGluThrGlnCysTyrValVala 290
877 TCTCTATGATGAGCGCGGCTTATACAGAGCGGCTGTCGCGGG 926
290 lAlaIaIaGlnCysGlyArgHisGlyLysArgAlaSerTyrGlyLys 306
927 CCGGCGGCAAGGTGGCATG...GAGGAGACTGATGCTCTCGCGCAC 973
307 SerMetValValaAspProTyrGlyTyrValValaAlaArgCysSerGlu.. 322
974 TCTGTCATGTCGGGCGGACCGGGAATGCTGCTCATACGACGCT 1023
323 GlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnL 339
1024 GGAAGACGAGGTGATCACCGCGCTGATCTGATCGCTCGCGGAAAC 1073
339 euATGATGHisLeuProValPhe...GlnHisArgArgProAspLeuTyr 354
1074 TGCCTGAAACATCTTCAACTTCAAGCAGCATGTCACGCCACGACCAT 1123
355 Gly 355

```

```

1124 GGT 1126
seq.name: /cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-876-398A-25
seq_documentation_block:
; Sequence 25, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKANO, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold G.
; REGISTRATION NUMBER: 25,258
; TELEPHONE: (202) 672-5300
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5399
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 pAD445
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-25

```

```

alignment_scores:
Quality: 214.50      Length: 401
Ratio: 1.089         Gaps: 17
Percent Similarity: 49.127   Percent Identity: 24.938

```

[illegible]

```

seq_name: /cgn2_6/plodata/2/ina/5A-COMB.seq:US-08-294-871A-5
seq_documentation_block:
  Sequence 5, Application US/08294871A
  Patent No. 5824522
  GENERAL INFORMATION:
    APPLICANT: Ikenaka, Yasuhiro
    APPLICANT: Namba, Hirokazu
    APPLICANT: Takano, Masayuki
    APPLICANT: Yajima, Kazuyoshi
    APPLICANT: Yamada, Yukio
    APPLICANT: Takahashi, Satomi
    APPLICANT: Okubo, Kazuma
    APPLICANT: Yamada, Kazuhiko
    APPLICANT: Hiraiishi, Yoshiro
    TITLE OF INVENTION: Immobilized Enzyme Preparation and
    TITLE OF INVENTION: Process for Producing D-a-Amino Acids
    NUMBER OF SEQUENCES: 70
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Foley & Lardner
      STREET: 3000 K Street, N.W.
      CITY: Washington
      STATE: D.C.
      COUNTRY: U.S.
      ZIP: 20007-5109
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/294,871A
      FILING DATE: 22-AUG-1994
    CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/971,758
      FILING DATE: 12-APR-1993
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/917,111
      FILING DATE: 07-AUG-1992
    PRIOR APPLICATION DATA:

```

```

1      APPLICATION NUMBER:  US 08/211,641
2      FILING DATE:  11-APR-1994
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER:  JP 400848/1990
5      FILING DATE:  07-DEC-1990
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER:  PCT/JP91/01696
8      FILING DATE:  06-DEC-1991
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER:  JP 407922/1990
11     FILING DATE:  27-DEC-1990
12     PRIOR APPLICATION DATA:
13     APPLICATION NUMBER:  JP 078840/1991
14     FILING DATE:  11-APR-1991
15     PRIOR APPLICATION DATA:
16     APPLICATION NUMBER:  JP 140051/1991
17     FILING DATE:  12-JUN-1991
18     PRIOR APPLICATION DATA:
19     APPLICATION NUMBER:  PCT/JP92/00739
20     FILING DATE:  10-JUN-1992
21     PRIOR APPLICATION DATA:
22     APPLICATION NUMBER:  JP 212692/1992
23     FILING DATE:  10-AUG-1992
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER:  PCT/JP93/01101
26     FILING DATE:  05-AUG-1993
27     PRIOR APPLICATION DATA:
28     APPLICATION NUMBER:  JP 340078/1992
29     FILING DATE:  21-DEC-1992
30     ATTORNEY/AGENT INFORMATION:
31     NAME:  Wegner, Harold C.
32     REGISTRATION NUMBER:  25, 258
33     REFERENCE/DOCKET NUMBER:  74129/127/AOPB
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE:  (202) 672-5300
36     TELEFAX:  (202) 672-5399
37     TELEX:  904136
38     INFORMATION FOR SEQ ID NO:  5:
39     SEQUENCE CHARACTERISTICS:
40     LENGTH:  1785 base pairs
41     TYPE:  nucleic acid
42     STRANDEDNESS:  double
43     TOPOLOGY:  linear
44     ORIGINAL SOURCE:
45     STRAIN:  JM109 PAD404 (FERM BP-3913)
46     FEATURE:
47     NAME/KEY:  CDS
48     LOCATION:  233..1144
49     OS-08-294-871A-5

```

```
alignment_scores:
  Quality: 213.50
  Ratio: 1.084
  Percent Similarity: 48.522
  Length: 406
  Gaps: 17
  Percent Identity: 24.384
```

alignment_block:

Align seg 1/1 to: US-08-294-871A-5 from: 1 to: 1785

[illegible]

40 lIePhr.....AngProPtoHIsNr 46
116 ACAACGTTTTCCGGCGGCTGGCGGACATCACCTAGAGAAAGACAG 225
46 gPhLeuSerLeuLeuCySProGlyLeuArgl eProGlnLeuSerVal 63
226 GTTCATATACA.....CGTAGATGATACTTG 251
63 euCSySlaGlnProArgProAlaMetAlaIleSerSerSerCyS 79
252 CAGTGGGACAACAGAGTCCGATCGCGCGCGGAGACACCGGACAGGTC 301
80 GluLeuProLeuValAlaValCySLeuValThrSerThrProAspLys 96
302 GTCTGTGTGTCTTCGACATG.....CTGACGAAGCGCGGAGCGGG 345
96 nGlnAspPheLeuThrcySAlaGluLeuValArgGlnAlaAlaArgLeu 113
346 CGCGAATTTCTATTGTCTTCCCGAATC..... 373
113 lYalAsySLeuAlaLeuLeuProGlnAlaPheAspPheIleAlaArgAsp 129
374 ..GGCCTTAACACCTCTTCCCGGC...TGCATTTTACACGACGAGGC 418
130 ProAlaGluThrLeuHisLeuSerGluProLeuGluGlyLysLeuLeu 146
419 GAGCTGCATAGCTTCTATGACAGCGAA...ATGCCCGCGCGCGTGTCCG 465
146 uGluTyrThrGlnLeuAlaArgGlnCySgLyLeuThrPheLeuSerLeuGly 163
466 TTCACCTCTTTAGAAAGCGCGGAGCTCGGAGTCCGCTCAATCTGGGCT 515
163 lYrPheHis.....GluArgGlyGlnAspTyrGluGlnThrGlnLys 176
516 ACAGTGAACCTCGTGTGCAAGCGGC.....GTCAAGCT 550
177 lLeTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaAlaTh 193
551 CGCTTCAACAGCTCCATTTTGATGATAACTCAGACAGATGCTGGCAA 600
193 rTyArgArgLysThrHisLeuCySAspValGluLeuProGlyGlnGlyProm 210
601 GTATCGAAGATTCATTGG.....CCGGTCCAGAG... 631
210 etCySgLeuSerAsnSerThrMetProGlyProSerLeuGluSerProVal 226
632GAGTACGAGGCGCTACCGCGCTTCACAGCATTTGAAAAAGCTTAT 676
227 SerThrPro.....AlaGlyLys 232
677 TTTCAGCGCGGCGATCTCGGCTCCCGGCTATGACGTGACGCGCGGAA 726
232 sIlleGlyLeuAlaValCySTyrAspMetArgPheProGluLeuSerLeuA 249
227 AATGGGATGTTCATCTGCAACAGATCGCGCGCTGTAAGCTGCGGG 776
249 lalaGlnAlaGlyLeuGluLeuThrTyrProSerAlaPheGly 265
777 TGATGGCGCTAAGGGCGCGCGAGATCATCTGC.....GGC 811
266 SerIleThrGlyProAlaHisTyrGluValLeu..... 276
812 GGTAAACAACGCGGACCCACAAATCCCTGTTCGCCACGACGACCACT 861
277LeuArgAlaArgAlaIleGluThrG 285
862 GACGCTCTTCACCAATCTCATCGATCAGCGCGGCTTTATCAGAACG 911
285 lncTyTyValValAlaAlaAlaGlnCySgLYrArgHisHisGlyLysArg 301
912 GGGCGCTGTCCGCGCGCGCGGCAAGGTGGCATG...GAGAGAACTCG 958
302 AlaSerTyrgLysHisSerMetValValAspProTyrGlyThrValValAl 318

```

959  ATGCTCTCGGGACACATCTCTGCATCTGTGGCCCGACGCGGGAAATCTGCG 1008
318  aATgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLeuA 334
1009  TCTCACTACGACGACCTGGAMGACGACGAGGTGACCGACCGCGCGTGCATGTCG 1058
1059  ATGCGTCCGGGAGACTGCTGGAAACACATCTTCAACTTCAAGCAGCATCTGT 1108

seq_name: /cogn2_6/ptodata/2/ina/5A-COMB.seq:US-08-294-871A-59

seq_documentation_block:
Sequence 59, Application US/08294871A
Patent No. 582452
GENERAL INFORMATION:
APPLICANT: Ikenaka, Yasuhiro
APPLICANT: Nanba, Hirokazu
APPLICANT: Takano, Masayuki
APPLICANT: Yajima, Kazuyoshi
APPLICANT: Yamada, Yukio
APPLICANT: Takahashi, Satomi
APPLICANT: Okubo, Kazuma
APPLICANT: Yamada, Kazuhiko
APPLICANT: Hiratschi, Yoshiro
TITLE OF INVENTION: Immobilized Enzyme Preparation and
Process for Producing D-a-Amino Acid
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,871A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,758
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,111
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/211,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 400848/1990
FILING DATE: 07-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP91/01696
FILING DATE: 06-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 407922/1990
FILING DATE: 27-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 078840/1991
FILING DATE: 11-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 140051/1991
FILING DATE: 12-JUN-1991

```

```

1      PRIOR APPLICATION DATA :
2      APPLICATION NUMBER: PCT/JP92/00739
3      FILING DATE: 10-JUN-1992
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: JP 212692/1992
6      FILING DATE: 10-AUG-1992
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: PCT/JP93/01101
9      FILING DATE: 05-AUG-1993
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: JP 340078/1992
12     FILING DATE: 21-DEC-1992
13     ATTORNEY/AGENT INFORMATION:
14     NAME: Wegner, Harold C.
15     REGISTRATION NUMBER: 25, 258
16     REFERENCE/DOCKET NUMBER: 74129/127/AOBA
17     TELECOMMUNICATION INFORMATION:
18     TELEPHONE: (202) 672-5300
19     TELEFAX: (202) 672-5399
20     TELEX: 904136
21     INFORMATION FOR SEO ID NO: 59:
22     SEQUENCE CHARACTERISTICS:
23     LENGTH: 1785 base pairs
24     TYPE: nucleic acid
25     STRANDEDNESS: double
26     TOPOLOGY: linear
27     ORIGINAL SOURCE:
28     STRAIN: JM109 ptd456
29     FEATURE:
30     NAME/KEY: CDS
31     LOCATION: 233..1144
32     OS-08-294-871A-59

```

alignment_scores:	
Quality:	213.50
Ratio:	1.084
Percent Similarity:	48.522
	Length:
	Gaps:
	Percent Identity:
	406
	17
	24.384

alignement_block:
US-09-357-675C-25 x US-08-294-871A-59

Align seg 1/1 to: US-08-294-871A-59 from: 1 to: 1785

2 LeuAlaAlaAlaA***LeuAlaIProAspArgProProAspArgThrLeuAr 18
 |||:::|||||::::::|||||:::|
 26 TTGTCAAGCAGGCGCAAAATTCGCGTTCGCGCTCGGTTGCACATCAAAAT 75
 18 gmeVAlIleuAlaIleSerSerCys..... 26
 ::|||:::||||||||
 76 TTTTACGCGCTTAAATGTCGTCGTCGATGTAATTAATTCGTACTTAAATGAG 125
 27ArgThrTyrSerLeuSerArgArgProArgLeuGlyPhe 39
 |||:::|||||:::|
 126 AATTTCATGTTGGCGCGCAGTCAAAACCGCGTTTGGCGGATGTGTTCC 175
 40 IleThr.....ArgProProHisAr 46
 |||:::|
 176 ACAACGTTTTCCCGCGCGCTGGCGCGGACATCACCTAGGAAGGACGACAG 225
 46 gPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 63
 |||:::|||||:::|
 226 GTTTCATGACA.....CGTCACATGATGACTTGG 251
 63 eucYsAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCys 79
 ::|||:::|||||:::|
 252 CAGTGGGACACAAAGTCCGATCGCGCGCGGAGACACCGGAAACAGATC 301
 80 GluLeuProLeuValAlaValaCysGlnValThrSerThrProAspLysG1 96
 ::|||::::::|
 302 GTCGTTTCGTCCTTCGACATC.....CTGACGAAGCGCGGACCGCGGG 345
 96 glnAsnAspPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuG 113

```

346 CGGAAATTCATGTCCTCCCGCACTC..... 373
113 lYAlaCySleuAlaPheLeuProGluAlaPheAspHeIleAlaArgAsp 129
374 ..GGCTTACAGACTTCTTCCCGCGGTAT...TTCACCGAGAGGCC 418
130 ProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuLeuGlu 146
419 GAGCTGATAGCTTCTATGAGACGAA...ATGCCGCGCCGGGTCG 465
146 uGluTyrThrGluLeuAlaArgGluGlyLeuTyrLeuSerLeuGly 163
466 TCACACTTGTGAGAGCGCGCAACTCGGATCGGCTTCAATCTGGGCT 515
163 lYPheHis.....GluArgGlyGlnAspTrrpGluThrGlnLys 176
516 ACCTGAACTCGTCGTGAAGCGGC.....GTCAAGCGT 550
177 lIeTyrAsnCySHisValLeuLeuAsnSerLysGlyAlaValAlaThr 193
551 CGCTTCAACACGTCATTTGGTGATAGTCAGGCAAGATCGTGGCAA 600
193 rTyrArgLysThrHisLeuCySAspValGluLeuProGlyGlnGlyPro 210
601 GATTCGTAAAGATTCATTG.....CCGGTCACAAAG... 631
210 eTcysGluSerAsnSerThrMetProGlyProSerLeuGluSerProVal 226
632 ....GAGTACGAGCGCTACCGCGCTTCACGATCTTGAAGACGTTAT 676
227 SerThrPro.....AlaGlyLys 232
677 TTCGACCGGGCGATCTCGGCTTCCGGTCTATGACGTCAAGCGGAA 726
232 sIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeu 249
727 AATGGGATGTTCATCTGCAGACATCGCGCTGCGTGAAGCGTGGGG 776
249 lAlaValAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 265
777 TGATGGCGCTCAGGCGCGGAGATCATCTGC.....GGC 811
266 SerIleThrGlyProAlaHisTrrpGluValLeu..... 276
812 GGCTACACAGCGCGACCCACAATCCCTGTTCGCCAGCAGCACACT 861
277 .....LeuArgAlaArgAlaIleGluThrG 285
862 GAGTCCTTCCACCATCTCTATCGATGCGAGCGGCTCTATCAGAACG 911
285 lncysTyrValValAlaAlaGlnCysGlyArgHisSHsGluLysArg 301
912 GGGCTGTGTCGCGCGCGGCAAGTCAGGCAATG...GAGGAGAACTGC 958
302 AlAserTyrGlyHisSerMetValAlaAspTrrpGlyThrValAla 318
959 ATCTGCTGCGCCACTCTGCACTGCGCGCACCGGAAACGTCGCGC 1008
318 aArgCysSerGlu...GlyProGlyLeuCySLeuAlaArgIleAspLeu 334
1009 TTCCTACTAGACGCTGGAAGACAGATGATCACCGCGCGTCACTCG 1058
334 snTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisArg 349
1059 ATGCGTCCGCGGAAGTCGTGAACACATCTTCAACTTCAAGCACACTCGT 1108
350 ArgProAspLeuTyrGly 355
1109 CAGCCCGACGACATATGCT 1126

```

seq_name: /sgn2_6/ptodata/2/lna/6A_COMB.seq:us-08-876-398A-5

```

seq_documentation_block:
; Sequence 5, Application US/08876398A
; Patent No. 6083752
;
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Megner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD404 (FERM BP-3913)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-5

```

alignment_scores:

Quality:	213.50	Length:	406
Ratio:	1.064	Gaps:	17
Percent Similarity:	48.522	Percent Identity:	24.384

alignment_block:

US-09-357-675C-25 x US-08-876-398A-5 ..

Align seg 1/1 to: US-08-876-398A-5 from: 1 to: 1785


```

466 TCACCTCTTGAAGAAGCCGCGGAACCTCGGGATCGGCTTCATCTATCTGGGCT 515
163 lYPheNis.....GluArgGlyGlnAspTrpGluGlnThrGlnLys 176
516 ACGCTGAACCTCGCGTCGCGAAGCGGGC.....GTCAACGCT 550
177 lIeTYAsnCysHisValLeuLeuAsnSerLysGlyValValAlaIatH 193
551 CGCTTCACACGCTCATTTGGTGGAATAACTCAAGCAAGATCGTCGGCAA 600
193 rTYArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProm 210
601 GATATCGTAAATCCATTTC.....CCGGTCACAGAAG.... 631
210 eTCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProVal 226
632 ....GAGTAACGAGGCTACCGGCGTTCAGCATCTTGAAACCGTTAT 676
227 SerThrPro.....AlaGly 232
677 TTCGACCCGGGCGATCTCGGCTTCCGGTCTATGACGTGACGCGCGAA 726
232 sIleGlyLeuAlaValLysrTyAspMetArgPheProGluLeuSerLeuA 249
727 AATGGGGAATTTATCTGCAACAGATCGCGCTGGCTGTGAACCTCGGGGG 776
249 lAlenuAaGlnAlaGlyAlaGluIleLeuThrTYrProSerAlaPheGly 265
777 TGAATGGGCTTCAGAGCGGCGCGAGATCATCTGC.....GGC 811
266 SerIleThrGlyProAlaHisTrpGluValLeu..... 276
812 GGCTACACACGCGCGACACCACAATCCCTGTTCGCCGACGACGACACT 861
277 .....LeuArgAlaArgAlaIleGlyThrG 285
862 GACGCTCTTCCACCACATCTCTATGAGATGCAAGCCGGGCTTATCAGAAAC 911
285 lncSTYrValValAlaAlaAlaGlnCysGlyArgHisGlnLysArg 301
912 GGGCTGATGTCGGGGCGCGGCGCAAGTCAGGACATG...GAGGAAGATCGC 958
302 AlASerTYrGlyHisSerMetValValAspProTrpGlyThrValAlaI 318
959 ATGCTCTCGGCGCATCTCTCATCTGTGGCGCCGACCGGGGAATCGTCGC 1008
318 aArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgLysLeuA 334
1009 TCTCACTTACAGCCTGGAGACGAGGAGATACCCGCGCGCTCATCTCG 1058
334 snTYrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisArg 349
1059 ATGGCTGCGCGGAACGTGGTGAACACATCTTCAACTTCAGACACATCGT 1108
350 ArgProAspLeuTYrGly 355
1109 CAAGCCCAAGCATATGCT 1126

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-294-871A-17

seq_documentation_block:
; Sequence 17, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hirataishi, Yoshiro

```

```

; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; STRAIN: JM109 pad434
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: 233..1144
; US-08-294-871A-17
;
; alignment_scores:
;   Quality: 212.50      Length: 407
;   Ratio: 1.095        Gaps: 17
;   Percent Similarity: 47.666      Percent Identity: 24.816
;
; alignment block:
; US-09-357-675c-25 x US-08-294-871A-17
;
; Align seg 1/1 to: US-08-294-871A-17 from: 1 to: 1785
;
; 2 LeuAlaIaLaLa**LeuAlaProAspArgProAspArgThrLeuAr 18
; |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; 26 TTGTCAAGCAGCGCAATTCGGTCCGCTCCGGTTACAGATCAAAAT 75
;
; 18 gMetValLeuAlaIleSerSerCys..... 26
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 76 TTACGCCCTGTATTCGTGTCGATGTAATTCGACTTATGTAG 125
;
; 27 .....ArgThrSerLeuSerArgArgProArgLeuGlyPhe 39
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 126 MATTGCATTGCCCGCGAGTCACAAAGCGGTTTCGGCGAGTCTTC 175
;
; 40 IleThr.....ArgProHisAr 46
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 176 ACAACGTTTCCCGCGCGTGGCCGACATCACCTAGGAAGAGCAGAG 225
;
; 46 gPheLeuSerLeuLeuCysProGlyLeuArgIleProGlyLeuVal 63
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 226 GTTCATGACA.....CGTCAGATGATCTTG 251
;
; 63 eucysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCys 79
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 252 CAGTGGACAAAGATCGATCGCGCGGAGACACCGCAACAGTGC 301
;
; 80 GluLeuProLeuValAlaValCysGlnValThrSerThrProAspGly 96
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 302 GTGCTTCGTTCTCTCGACATG....CTGACGAAGCCCGACCGGG 345
;
; 96 ngIAsnPheLeuThrcysAlaGluLeuValArgValAlaIaArgLeu 113
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 346 CGCGAATTTCATTGCTCTCCCGAATC..... 373
;
; 113 LysIaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArg 129
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 374 ..GCGCTTACGACCTTCTCCGCGC.....TGGTATTTCACCGAC 412
;
; 130 ProAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLeuLe 145
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 413 GAGCGCGAGTCGATAGCTTCTATGAGACCGAAATGCCCGCGGTGT 462
;
; 145 uGluGluTyrThrGlnLeuAlaArgGlyCysGlyLeuTyrLeuSerLeu 162
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 463 CCGTCACACTTGTGAGAAAGCCCGGAATCGGGATCGCTCAATCTGG 512
;
; 162 LysIlePheHis.....GluArgGlyLysPrrGluGlnThrGln 175
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 513 GCTACGCTGAATCGTCTGGAAGCGCGC.....GTCAGG 547
;
; 176 LysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 192
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 548 CGTCGCTTCAACAGTCATTTTGGTGATGAATCAGGCAAGATCGTCGG 597
;
; 192 atThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGly 209
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 598 CAAGTATCGTAAGATCCATTG.....CCGGTTCACAAAG. 631
;
; 209 rometCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 225
; |||:|||||:|||||:|||||:|||||:|||||:|||||:
; 632 .....GAGTACGAGGCGCTTACCGCGCTTCCAGCATCTTGAAAGCGT 673

```


[illegible]

```

      FILING DATE: 22-AUG-1994
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/971,758
      FILING DATE: 12-APR-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/917,111
      FILING DATE: 07-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/211,641
      FILING DATE: 11-APR-1994
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 400848/1990
      FILING DATE: 07-DEC-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP91/01696
      FILING DATE: 06-DEC-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 407922/1990
      FILING DATE: 27-DEC-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 078840/1991
      FILING DATE: 11-APR-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 140051/1991
      FILING DATE: 12-JUN-1991
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP92/00739
      FILING DATE: 10-JUN-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 212692/1992
      FILING DATE: 10-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/JP93/01101
      FILING DATE: 05-AUG-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: JP 340078/1992
      FILING DATE: 21-DEC-1992
      ATTORNEY/AGENT INFORMATION:
      NAME: Wegner, Harold C.
      REGISTRATION NUMBER: 25,258
      REFERENCE/DOCKET NUMBER: 74129/127/NOA
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202) 672-5300
      TELEFAX: (202) 672-5399
      TELEX: 904136
      INFORMATION FOR SEQ ID NO: 19:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1785 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      ORIGINAL SOURCE:
      STRAIN: JM109 PAD435
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 233..1144
      US-08-294-871A-19

alignment_scores:
      Quality: 212.50      Length: 407
      Ratio: 1.095      Gaps: 17
      Percent Similarity: 47.666      Percent Identity: 24.816

alignment_block:
      US-09-357-675C-25 x US-08-294-871A-19      ..

Align seg 1/1 to: US-08-294-871A-19 from: 1 to: 1785

2 LeuAlaAlaAla***LeuAlaProAspArgProProAspArgThrLeuAr 18
|||||:::||||:::|||||:::
26 TTGTCAAGCAGGCGAATTCGGTTCCGCTCGGTTGACAGATCAAAAAT 75

```



```

231 ylystlleglyleuAlaValCysTyrrAspmMetArgPheProgluleuSertL 248
      :::::::::::::: ::::||| ||| ||:::|||||
724 GAATAGGGGATCTTCACTgCAACAGATCGCCGTGGCTGAAGCTGC 773
248 eualaleuAlaglnAlaGlyIleuIleuThrTyProSerAlaphe 264
      :::::::::::: ||||||::::::::::
774 GGGTGATGGGCCTCAGGGGGCCGAGATCATCTGC..... 808
265 GlySerIleThrGlyProAla..... 271
      ||::: :: |:::
809 GGCGGCTACAAcAGCGCACCAATACCCTGTTCGCCAGCAGCACCA 858
272 ..... HistripleuAlleuLeuArgAlaAlaIleGlut 284
      ||| ..::: :::::|::: :::::
859 CCTGAGCTCTTCACCATCTCTATGATCATCAGCCGGCTTTATACA 908
284 hrcInCysTyrrValAlaAlaAlaIncysglnArgHIsigIulyr 300
      :: :::::|:::|:::|:::|:::|:::
909 ACGGGGCTGTGTCGGCGCCGGCGGCAAGTGCGCATG... GAGAAGAc 955
301 ArgAlaSerTygIlyHisSerMetValAlaAspTrtArgIyThrValVa 317
      |||||||:::|::: ||| ||| ::::|
956 TGcATGCTGTCGGCACCTCTGATGCTGGCGCGAGCGGGAAATGCT 1005
317 lAlaArgCysSerGlu..GlyProGlyLeuCysLeuAlaArgIleAspL 333
      |||| :::::|::: |||
1006 CGCTTCACATCAcAGCGCTGGAAGAGAGAGTATCACCCGCGCTGCATC 1055
333 euAnsntyrLeuArgInleuArgArghIsleuProvalPhe.. GlNHs 348
      ||::: |:::|:::|:::|:::|:::|:::|:::|:::|
1056 TCgATGCTGCGCGGAGACTGCGTAACACATCTTCACTTCAAGCAGCAT 1105
349 ArgArgProAspLeuTyrgly 355
      ||:::|:::|::: |:::|:::
1106 CGrCACGCCCAcAGCATTTGgr 1126

seq_name: /cgm2_6/ptodata/2/lna/6A_COMB.seq:US-08-876-398A-19
seq_documentation_block:
Sequence 19, Application US/08876398A
Patent No. 6083752
GENERAL INFORMATION:
APPLICANT: IKENAKA, Yasuhiko
APPLICANT: NABANA, Hiroyasu
APPLICANT: TAKANO, Masayuki
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: YAMADA, Yukio
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN THERMOSTABILITY AND USE THEREOF
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
CITY: Washington
STREET: 3000 K Street, N.W.
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,398A
FILING DATE: 16-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/221,641
FILING DATE: 11-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP93/01101

```

```

1      FILING DATE: 05-AUG-1993
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: JP 340078/1992
4      FILING DATE: 21-DEC-1992
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: JP 212692/1992
7      FILING DATE: 10-AUG-1992
8      ATTORNEY/AGENT INFORMATION:
9      NAME: Wegner, Harold C.
10     REGISTRATION NUMBER: 25,258
11     REFERENCE/DOCKET NUMBER: 74129/130
12     TELECOMMUNICATION INFORMATION:
13     TELEPHONE: (202) 672-5300
14     TELEFAX: (202) 672-5399
15     INFORMATION FOR SEQ ID NO: 19:
16     SEQUENCE CHARACTERISTICS:
17     LENGTH: 1785 base pairs
18     TYPE: nucleic acid
19     STRANDEDNESS: double
20     TOPOLOGY: linear
21     MOLECULE TYPE: DNA (genomic)
22     ORIGINAL SOURCE:
23     ORGANISM: Escherichia coli
24     STRAIN: JM109 pAD35
25     FEATURE:
26     NAME/KEY: CDS
27     LOCATION: join(233..1141)
28     US-08-876-398A-19
29
30 alignment_scores:
31     Quality: 212.50      Length: 407
32     Ratio: 1.095      Gaps: 17
33     Percent Similarity: 47.666      Percent Identity: 24.816
34
35 alignment block:
36 US-09-357-675C-25 x US-08-876-398A-19 ..
37
38 Align seg 1/1 to: US-08-876-398A-19 from: 1 to: 1785
39
40 2 LeuAlaIaAla***LeuAlaProspaRgProPaSPaRgThreuar 18
41 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 26 TTGTCAAGCAGCGGAAATTCGCGTTCCGCTCCGTGACAGATCAAAAT 75
43 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44 18 gmetValLeuAlaIleSerSerCys..... 26
45 76 TTTAAGCGCTGTATTATGTCGTCTGATGATTAATTTCCGACTTATGTC 125
46 .....ArgThrTyrSerLeuSerArgArgProaRgLeuglyphe 39
47 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 126 AATTTCGATTGCGCCGCGAGTCAACAAACCGGTTTCGCGCATGTCTTC 175
49 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50 40 IleThr.....ArgProProHisar 46
51 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52 176 ACAACGTTTCCCGCGCTGGCCGGAATCACCTAGGAAGAGCAGAG 225
53 .....GpHeLeuSerLeuSerProGlyLeuArgTleProGlnLeuSerVal 63
54 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 226 GTTCATATACA.....CGTCGATGATCTTG 251
56 eucYsaIaGlnProaRgProaRgaIaMetAlaIleSerSerSerCys 79
57 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
58 252 CAGTGGACAAACAGATCCGATCGCGCGCGAGACACGCAACAGTTC 301
59 .....GlnLeuProLeuValAlaValCysGlnValTMrSerThnProaPlyScl 96
60 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 302 GTCGTTTCGCTTTCGCAATG.....CTGACGAAGAAGCCGACCGCGGG 345
62 .....ngInaSPheLeuThrcYsaIaGlnLeuValaRgIuaIaIaIaRgLeug 113
63 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 346 CGGGAATTTATTTCTTCCCGAATC..... 373
65 113 lyaIaCysLeuAlaPheLeuProGlnAlaPheASPheIleAlaIaRgasp 129

```



```

      ::|||::|||::|||::|||::|||
1059 ATCGCTCCGGGAACGCGTGAACACATCTTCAACTTCAAGCAGCATCGT 1108
      350 ArgProAspLeuTyrGly 355
      ::|||::|||::|||
1109 CAGGCCCGCAGCATATGGT 1126

```

THIS PAGE BLANK (USPTO)